HELSINKI UNIVERSITY OF TECHNOLOGY
Department of Computer Science
and Engineering

Jarkko Ylipaavalniemi

Variability of Independent Components in functional Magnetic Resonance Imaging

Master's thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Technology

Espoo, March 3, 2005

Supervisor: Academy Professor Erkki Oja
Instructor: Docent Ricardo Vigário

TEKNILLINEN KORKEAKOULU
TIETOTEKNIIKAN OSASTO

DIPLOMITYÖN TIIVISTELMÄ

Tekijä:	Jarkko Ylipaavalniemi
Osasto:	Tietotekniikan osasto
Pääaine:	Informaatiotekniikka
Sivuaine:	Teollisuustalous
Työn nimi:	Riippumattomien komponenttien variaatiot toiminnallisessa magneettikuvauksessa
Päiväys:	3.3.2005
Sivumäärä:	82
Professuuri:	T-61 Informaatiotekniikka
Työn valvoja:	Akatemiaprofessori Erkki Oja
Työn ohjaaja:	Dosentti Ricardo Vigário

Tiivistelmä:

Riippumattomien komponenttien analyysi (ICA) on laajalti käytetty ja tehokas datalähtöinen signaalinkäsittelymenetelmä. Vaikka se onkin osoittautunut hyödylliseksi monilla tutkimusaloilla, kuten bio- ja lääketieteessä, tietoliikenteessä, taloudessa ja luonnollisten kuvien käsittelyssä, liittyy sen käyttämiseen ongelmia. Eräs ongelma on, että ICA:n tuottamat tulokset vaihtelevat hieman jokaisen sovelluskerran välillä. Tästä johtuen tulosten luotettavuus on helppo kyseenalaistaa. Tulosten variaatiot ovat seurausta menetelmän käyttämän tiedon ja toteutuksen sisältämistä tilastollisista ominaisuuksista. Nämä huomiot koskevat myös monia muita sokean lähteiden erottelun (BSS) menetelmiä. Tämä työ esittelee menetelmän tulosten konsistenttisuuden tutkimiseen ja kuvaa, kuinka variaatioita voidaan käyttää hyväksi hyödyllisen lisätiedon saamiseksi. Menetelmä perustuu useiden ICA:n suorituskertojen tuottamien tulosten ryhmittelyyn. Menetelmää on kokeiltu oikean toiminnallisen magneettikuvauksen (fMRI) yhteydessä, jossa käytettiin ääniärsykkeitä. Menetelmä tunnistaa useat riippumattomat komponentit konsistenteiksi, mutta tarjoaa myös arvokasta tietoa vähemmän konsistenttien ilmiöiden ymmärtämiseksi.

Avainsanat:

variaatiot, riippumattomien komponenttien analyysi, ICA, toiminnallinen magneettikuvaus, fMRI, uudelleennäytteistys, ryhmittely

HELSINKI UNIVERSITY OF TECHNOLOGY
DEPARTMENT OF COMPUTER SCIENCE
AND ENGINEERING

ABSTRACT OF MASTER'S THESIS

Author:	Jarkko Ylipaavalniemi
Department:	Department of Computer Science and Engineering
Major subject:	Computer and Information Science
Minor subject:	Industrial Engineering and Management
Title of thesis:	Variability of Independent Components in functional Magnetic Resonance Imaging
Date:	3.3.2005
Pages:	82
Professorship:	T-61 Computer and Information Science
Supervisor:	Academy Professor Erkki Oja
Instructor:	Docent Ricardo Vigário

Abstract:

Independent component analysis (ICA) has been widely adopted as a powerful data-driven signal processing technique. But, even though it has been shown to be helpful in many fields, such as, biomedical systems, telecommunication, finance and natural image processing, there remains problems in its wide adoption. One concern is that solutions found with ICA algorithms tend to change slightly each time analysis is done, raising serious questions about the reliability of those solutions. This behavior stems from the stochastic nature of the data and ICA algorithms, and affects many other blind source separation (BSS) algorithms as well. This thesis presents a method to analyze the consistency of the solutions. It is also shown how to exploit the variability to gain additional information on the found solutions. The method is based on clustering solutions from multiple runs of bootstrapped ICA. Its usefulness is tested with a real functional magnetic resonance imaging (fMRI) experiment, involving auditory stimulus, where several independent components are truly consistent. Additionally, the information acquired with the method helps in analyzing the underlying phenomena of the less consistent ones.

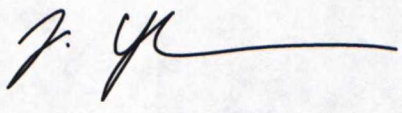
Keywords:

variability, independent component analysis, ICA, functional magnetic resonance imaging, fMRI, bootstrapping, clustering

Preface

This work was done at the Neural Networks Research Centre of the Laboratory of Computer and Information Science in Helsinki University of Technology during the year 2004. I am grateful to my supervisor Academy Professor Erkki Oja and instructor Docent Ricardo Vigário for giving me the possibility to work in the Neuroinformatics group and allowing me to research this interesting topic for my thesis. I am also grateful to the whole group and the staff of the laboratory for the discussions which gave me many ideas for this thesis and future research. I would also like to thank Academy Professor Riitta Hari from the Advanced Magnetic Imaging Centre (AMI-Centre) and the Brain Research Unit of the Low Temperature Laboratory for discussions and invaluable physical and medical insight. Finally, I wish to thank my family and friends for encouraging and supporting me in my studies.

Otaniemi, March 3, 2005



Jarkko Ylipaavalniemi

Contents

List of Figures

Symbols and Abbreviations

$t = 1, \ldots, T$	Time sequence with T observations.
$k = 1, \ldots, K$	Source sequence with K sources.
$v = 1, \ldots, V$	Volumetric sequence with V voxels.
$\tilde{\mathbf{X}}$	Observed data matrix.
$\mathbf{X}$	Whitened data matrix.
$\tilde{\mathbf{A}}$	Arbitrary mixing matrix.
$\mathbf{A}$	Orthonormal mixing matrix.
$\hat{\mathbf{A}}$	Normalized concatenated mixing matrix.
$\mathbf{W}$	Orthonormal demixing matrix.
$\mathbf{S}$	Independent sources matrix.
$\mathbf{x}_t^T$	Row vector of matrix $\mathbf{X}$.
$\mathbf{a}_k$	Column vector of matrix $\mathbf{A}$.
$\bar{\mathbf{a}}_k$	Mean value of vector $\mathbf{a}_k$.
$\mathbf{w}_k^T$	Row vector of matrix $\mathbf{W}$.
$\mathbf{s}_k^T$	Row vector of matrix $\mathbf{S}$.
a_{ij}	Scalar element of matrix $\mathbf{A}$.
$\mathbf{M}$	Arbitrary square matrix.
λ_i	The ith eigenvalue.
$\mathbf{v}_i$	Eigenvector corresponding to λ_i.
$\mathbf{D}$	Diagonal matrix of eigenvalues.
$\mathbf{V}$	Orthogonal matrix of eigenvectors.
$g(\cdot)$	Nonlinear function.
$g'(\cdot)$	Derivative of function $g(\cdot)$.
$E\{\cdot\}$	Statistical expectation function.
$p(\cdot)$	Probability density function.
$f(\cdot)$	Arbitrary function.
$\mathbf{C}$	Correlation matrix.
c_{ij}	Scalar element of matrix $\mathbf{C}$.
$\tilde{\mathbf{C}}$	Thresholded binary correlation matrix.
$\mathbf{R}$	Relations matrix.
r_{ij}	Scalar element of matrix $\mathbf{R}$.

CNS	Central Nervous System
EEG	ElectroEncephaloGraphy
MEG	MagnetoEncephaloGraphy
CT	Computed Tomography
PET	Positron Emission Tomography
MRI	Magnetic Resonance Imaging
fMRI	functional Magnetic Resonance Imaging
NIRS	Near-InfraRed Spectroscopy
DOI	Diffuse Optical Imaging
BOLD	Blood Oxygenation Level Dependent
GLM	General Linear Model
SPM	Statistical Parametric Mapping
BSS	Blind Source Separation
SOBI	Second-Order Blind Identification
TDSEP	Temporal Decorrelation source SEParation
PCA	Principal Component Analysis
EVD	EigenValue Decomposition
ICA	Independent Component Analysis
DSS	Denoising Source Separation

Glossary of Anatomical Terms

Superior	Above another part, opposite to inferior.
Inferior	Below another part, opposite to superior.
Anterior	Toward the front, opposite to posterior.
Posterior	Toward the back, opposite to anterior.
Lateral	Toward the side, away from the midline.
Medial	Toward the midline, away from the side.
Ipsilateral	On the same side of the body.
Contralateral	On the opposite side of the body.
Coronal plane	A plane seen from the front.
Sagittal plane	A plane seen from the side.
Horizontal plane	A plane seen from above.
Nucleus	A cluster of neuron cell bodies within the central nervous system. (Plural: nuclei)
Ganglion	A cluster of neuron cell bodies outside the central nervous system.
Gyrus	A protuberance on the surface of the brain. (Plural: gyri)
Sulcus	A fold or groove that separates one gyrus from another. (Plural: sulci)
Fissure	A long, deep sulcus

Chapter 1

Introduction

1.1 Background

Independent component analysis (ICA) is a recently introduced signal processing technique to solve the blind source separation (BSS) problem in a clear data-driven approach. BSS consists of finding underlying source signals from their observed linear mixtures. This is very difficult since both the mixing and sources are unknown. The solution of the BSS problem has useful applications in many research fields, including biomedical systems, telecommunication and finance. With ICA the problem is solved by assuming that the source signals are statistically independent, which has proven to be quite a natural assumption in many cases. ICA has been used to, for example, identify interesting signals, remove artifacts and reduce noise. ICA can also be used in feature extraction, having applications in, for example, natural image processing and human vision.

One very promising area for ICA is functional brain imaging. Using methods like functional magnetic resonance imaging (fMRI), it is possible to record signals related to the neural activity of the human brain. However, the complex and unpredictable nature of such data make it difficult to analyze using classical signal processing methods, for example, based on modeling the signals. ICA has been recently used with great success in several fMRI studies and may offer the possibility to conduct more advanced studies than before.

Yet, some problems remain in the wide adoption of ICA. One concern is the fact that the solutions found can change slightly each time ICA is applied, naturally causing one to question the reliability of the method. This differing

nature of the solutions can be caused by many factors. For example, the strict assumption of statistical independence may not hold for the data, or the estimation process in the ICA algorithm may be inherently stochastic. Also, additive noise can make even the most robust algorithm find varying solutions. Thus, it is difficult to know how stable and reliable the estimated solutions are.

Usually ICA solutions have been compared to those obtained with other methods, or an expert has estimated their feasibility under current knowledge. Needless to say, this effectively cancels the benefits attainable with strictly data-driven analysis. Moreover, such an expert might not even exist. On the other hand, bootstrapping has been successful in identifying consistent solutions. Bootstrapping means controlled resampling of the data and, essentially, allows the statistical analysis of the behavior of an algorithm. It has even been used to group consistently appearing solutions. However, the potential of bootstrapping and analyzing the consistency has not yet been fully exploited.

Increased interest on the matter has convinced that it is important to develop an efficient and usable method to better analyze the consistency of the solutions and characterize the affecting phenomena.

1.2 Purpose of the Work

The purpose was to develop an efficient method to exploit the variability of independent components based on existing tools and earlier experiments of the research group (most recently in Ylipaavalniemi and Vigário, 2004) as an improved alternative to other methods recently developed.

The method is based on running ICA multiple times in a bootstrapping manner and then clustering the solutions. Moreover, the method actually exploits the inherent stochastic variability to improve the solutions and to gain further information, which helps to properly interpret the solutions, even without an expert.

The usefulness was tested with real functional magnetic resonance imaging (fMRI) data, which uses a speech stimulus. There the method reveals interesting components with consistent stimulus-related activation patterns. Additionally, the results include other interesting components, whose time-courses are only mildly related to the stimulus, hence difficult to detect with traditional methods. Moreover, the experiments reveal also less consistent, yet interesting phenomena, which are hard to interpret using other methods.

An additional goal was to develop the needed tools to process and visualize fMRI data. The visualization tools are important during the interpretation of the results, and may eventually be published as an easy to use toolbox.

1.3 Structure of the Thesis

The thesis begins by introducing the basics of functional brain imaging in Chapter 2 and also briefly explaining the traditional analysis method. Chapter 3 is an introduction to independent component analysis and the chosen algorithm implementation. It also explains the theoretical and algorithmic problems in utilizing such an algorithm. Finally, it describes how ICA can be applied to fMRI data.

Chapter 4 considers the actual variability of independent components and presents the method to exploit it in analysis. After that, Chapter 5 explains how the analysis results can be visualized to allow easy and correct interpretation.

The experiments to test the method are described in Chapter 6 and the results are presented in Chapter 7. Finally, Chapter 8 gives the conclusions drawn from the work.

The thesis ends with Appendix A giving further information on the key mathematical concepts and Appendix B showing the complete results of the experiments.

Chapter 2

Brain Imaging

2.1 The Human Brain

The human brain has been one of the biggest research topics in many sciences, including biomedical, psychology and information theory. Current knowledge of the structure and function of the brain is substantial and growing fast, due to new imaging and analysis techniques. Consider reading Kalat (2003) for a great overview on the human brain. The key concepts needed to understand the thesis are explained next.

2.1.1 Structural Anatomy

The basic anatomical structure of the human brain is depicted in Figure 2.1. It has been studied for a long time, even before there were any ideas on what the brain does or how it functions. Much of the knowledge is based on histological studies, but under extreme cases, the living brain has been studied with crude and invasive methods. Currently, less invasive imaging methods allow the study of the structure and changes in it during the life of a subject, for instance, under a progressing illness.

The central nervous system (CNS) is formed by the cortex, brain stem, cerebellum and other connected subcortical areas. The brain stem and other subcortical regions are mainly involved in lower level functions, like automation and primitive signal processing. Higher functions, such as conscious thought, are performed on the cortex, that is, the surface of the brain. Most higher functions, like memory, also rely on support from the subcortical areas.

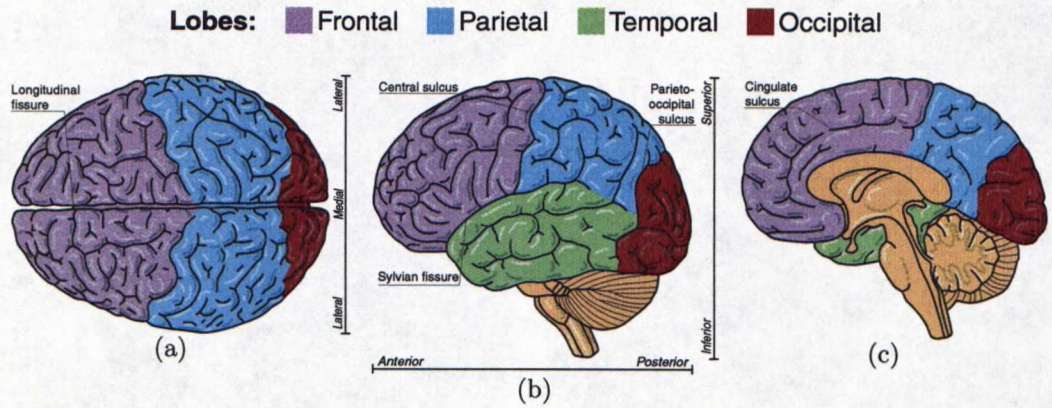


Figure 2.1: Anatomical structure of the human brain. (a) The horizontal view from above and the sagittal views from (b) the side and (c) middle of the brain show the basic structures and divisions, including the four lobes separated by sulci and fissures. Important names and directions are also shown.

There are essentially two kinds of tissues, the gray matter and the white matter. The gray matter contains the actual cell bodies of the neurons and most of it is concentrated on the surface, or cortex, of the brain. The connecting structures between the neurons form the white matter. The inside of the brain is mostly white matter, but there are some nuclei, which are small groups of neurons and basically work as junctions along signaling pathways.

The surface is heavily folded to increase its area. The folds are called sulci and separate the surface into small sections, or gyri. Bigger folds that separate larger parts are called fissures, like the longitudinal fissure in Figure 2.1(a), which separates the left and right hemispheres of the brain.

The division of the cortex into the four lobes, as shown in Figure 2.1(b), may seem somewhat arbitrary, but is based on major sulci and fissures, visible on the surface. Additionally, fine details, like the density of neurons and their size and shape, differ between the areas. Naturally, the boundaries are not always so clear in real brains, and can change slightly from one individual to another.

2.1.2 Functional Anatomy

As the gray matter is mostly on the surface of the brain, functional anatomy mainly details areas on the surface. But the connections are also very important. The neuronal configuration is similar throughout the surface, but different inputs and outputs of the peripheral nervous system are connected to

different parts of the brain. Thus, different areas of the brain are involved with different kind of information and serve a different purpose. Table 2.1 provides a quick lookup of some of the main details.

Lobe	Input and Output	Other Functions
Frontal	Motor	Memory and Emotions
Temporal	Auditory	Language and Structure
Parietal	Somatosensory	Association and Attention
Occipital	Visual	Pattern and Object Recognition

Table 2.1: Functional properties of the four lobes of the brain.

Figure 2.2(a) shows the location of some of the well known primary processing areas on the cortex more accurately. These areas are mainly connected contralaterally, which means that the areas on the left hemisphere are mainly responsible for signals from the right side of the body. The primary areas are then connected to additional areas nearby on the same hemisphere, or ipsilaterally. The additional areas usually perform more complex functions based on the processing done on the primary areas. The left and right hemispheres of the brain are functionally quite symmetric, but usually each task has a more dominant side. The brain is also adaptive in the sense that sometimes other areas overtake more functionality, when the dominant side suffers an injury.

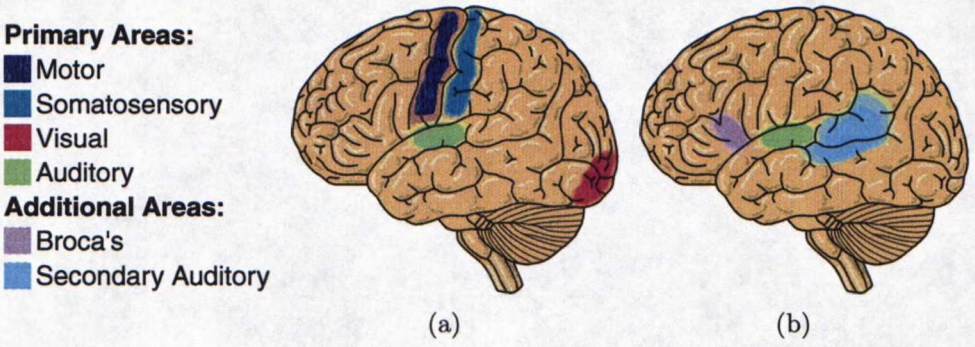


Figure 2.2: Functional areas of the human brain. (a) Primary areas involved in processing different information and (b) areas involved in auditory and speech signal processing.

2.1.3 Auditory Signal Processing

At first, audio processing may seem relatively simple compared to, for example, visual image processing. However, auditory processing is closely linked to

understanding spoken language and therefore also to higher functions, such as memory and conscious thinking. Audio processing in the brain has been studied extensively and still remains an interesting topic in current research.

The signal processing actually begins already in the ear and in the thalamus, even before the signals reach the cortex. The early processing is used to form a tonotopic map, based on frequency, on the primary auditory area. Also other processes, like focusing attention, affects how signals arrive at the cortex. Unlike in many other sensory inputs, audio signals from both ears are used together, which makes it possible to detect the direction of the original sound by analyzing the phases of the signals.

The primary auditory area responds to all kinds of sounds, but it is tightly connected with additional areas involved in more complex processing. These areas are shown in Figure 2.2(b) and are involved in, for example, understanding speech and forming sentences. The addional areas are usually more active on the left hemisphere. The precise function of these and other areas is currently studied (*c.f.*, Calhoun et al., 2001a, Bartels and Zeki, 2004, Arnott et al., 2004) and much remains to be discovered.

2.2 Evolution of Imaging Techniques

For a long time, only crude pathological methods existed for studying the brain, and functional studies were virtually impossible. Brain imaging has been developing rapidly during the last decades. Specifically in recent years, it has become also quite noninvasive, allowing the routine imaging of living tissues. For example, Bankman (2000) describes many recent brain imaging techniques and how they are utilized in the biomedical field.

Histological studies, using small tissue samples from pathological analysis, formed the basis of understanding, and motivated the development of methods to study the living brain. The oldest imaging methods use the gamma radiation generated by radioactive decay, commonly known as x-rays, to form images based on differing absorption properties of tissues. Taking x-ray images is very quick and simple, much like using an ordinary pocket camera. But repeated exposure to the radiation can be harmful and the contrast in soft tissues is not very good. Most of us are painfully aware how x-rays are still very widely used, for example, to image bone fractures.

For many years, studying brain functions was possible only by observing the relations between different areas of the brain and different functions when com-

paring the performance of healthy subjects to that of brain-damaged subjects. For example, aphasia refers to brain-damage that affects language skills.

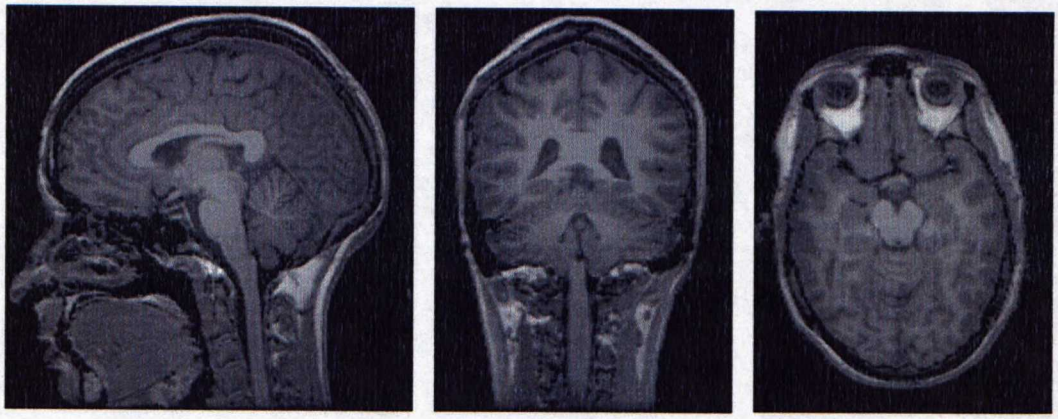
Electroencephalography (EEG) and magnetoencephalography (MEG) (*c.f.*, Niedermeyer and da Silva, 2004) changed the situation by allowing direct measurements of brain activity. EEG is based on measuring small changes in the electric field caused by neuronal activity, using small sensors attached to the scalp. In MEG, superconducting sensors positioned close to the scalp are used to measure changes in the related magnetic field. EEG and MEG are perhaps the most widely used techniques for functional studies of the brain, such as evoked responses, because they are completely noninvasive and offer very high temporal resolution. However, the inverse problem of finding out the originating location of the brain-waves accurately is very difficult. In some contexts, EEG and MEG are not considered as true imaging methods.

Advances in computer technology made computed tomography (CT) (Andreasen, 1988) possible. Essentially, the idea in CT is to take several x-ray images of a target from many directions and, using a computer, carefully combine the information in each two-dimensional image to form a single three-dimensional image of the target. CT can produce high resolution images, but suffers from the same problems as ordinary x-ray images, and is mainly usable for structural imaging. Additionally, taking many images naturally multiplies the exposure of the subject to the potentially harmful radiation.

Positron emission tomography (PET) (Andreasen, 1988) also produces three-dimensional images, but is based on radioactive markers injected to or inhaled by the subject. The marker substances can be pharmacologically designed to participate in certain metabolic reactions. The fast decaying also improves the image contrast. PET can produce images targeted at specific organs or processes related to illnesses, such as cancer, but the temporal resolution is not very good due to the slowness of the metabolic changes. Additionally, injecting radioactive substances into the body is invasive and potentially harmful. Furthermore, producing the needed beta-decaying substances is very difficult and expensive. Their rapid decaying also makes their usage difficult. Thus, PET is mostly used in studying severe illnesses.

Magnetic resonance imaging (MRI) (Moonen et al., 1990) is a virtually noninvasive method based on nuclear magnetic resonance, which is rather complex and shortly explained in Section 2.3. Because of the noninvasive nature and ability to produce high quality images, MRI has quickly become a popular method. Examples of such MR-images are shown in Figure 2.3. They are scans of a human head, using a setup that produces good contrast between different tissue types, thus revealing the anatomical structure in great detail.

The method is also very suitable for functional imaging, allowing a good compromise between spatial and temporal resolution.



(a) (b) (c)

Figure 2.3: Examples of structural magnetic resonance images. The images show slices of a human head viewed from (a) sagittal, (b) frontal and (c) horizontal directions.

Recently, near-infrared spectroscopy (NIRS) or diffuse optical imaging (DOI) (Meek et al., 1995) was proposed as a potential future technique. It is based on the diffusion of laser-light in tissue and blood. Laser based imaging would be fast and not sensitive to electromagnetic interference, allowing much cheaper imaging. The difficulties in using the method include the ability to generate high resolution images and penetrate deep enough into the tissues. In time, NIRS may become yet another widely used imaging technique.

Practically all the methods suffer from interference caused by the surrounding environment, usually in the form of electromagnetic radiation, but each technique has its benefits and may suit certain tasks better than the others. Newer methods are not always developed to replace the older ones, but may rather aim at solving some specific problems. All the methods mentioned before are still in use, and being improved to make them faster or less invasive. The rest of the thesis, however, focuses on the functional form of MRI, which is explained next.

2.3 Functional Magnetic Resonance Imaging

The complex physics of nuclear magnetic phenomena are very interesting, but certainly beyond the scope of this thesis. Therefore, only an overview of the

main concepts is given in this section, which also covers the standard processing and analysis methods used in fMRI studies. Consider reading Huettel et al. (2004) for a more thorough introduction.

Nuclear particles, such as protons and neutrons, have a magnetic property called *spin*, which behaves much like an ordinary dipole-magnet. Usually, the focus of interest is on hydrogen nuclei, because of their abundance in all tissue types and relatively simple spin behavior. The behavior of the nuclear spin of heavier atoms is more complicated, since there are internal interactions between the individual particles. Fundamentally, all MR-imaging is based on the interaction between the imaged tissue, externally applied magnetic fields and carefully synchronized radio frequency pulses.

Under a very strong and uniform magnetic field, the spins try to align parallel to the field either in the same or opposite directions. More precisely, the spins precess around those directions in a minimum energy state. Although the precession is not coherent, all the spins have a characteristic resonance frequency, proportional to the strength of the magnetic field. When a radio pulse in the resonance frequency is emitted, the spins absorb the energy and are forced into coherent precession. After the pulse, the absorbed energy decays in a relaxation process. The process is actually quite complex, for example, due to possible internal interactions in the tissue. However, signals measuring the decay of energy make the imaging possible, and adjusting the properties and timing of the radio pulses produces signals related to different aspects of the relaxation process.

To produce a volumetric (3-dimensional) image, the behavior of the spins is controlled more precisely with two gradient magnetic fields, which are perpendicular to each other. The first is applied in the same direction as the uniform magnetic field, causing the total strength of the field to change slightly along that direction. This makes the resonance frequency of the spins also different along its axis. The other gradient field is similar, but turned on and off repeatedly. As the spins precess faster during the application of the field, the spins along its axis accumulate different phases. The gradient magnetic fields allow focusing on a planar slice, which is defined by the axes of frequency-phase space. Carefully synchronizing the radio pulses with the fields produces signals originating from different parts of the slice. Additionally, the thickness of the slice can be controlled with the bandwidth of the radio pulses.

Using standard signal processing techniques, the measurements can be turned into an image of the focused slice. The full volume is produced by scanning several adjacent slices, one after the other. The image *voxels* contain a kind of density measure based on the scanning parameters and the properties of the

tissue. For example, with certain parameters the image is directly related to the proton density of the tissue.

The scanning is actually very slow since the relaxation process, and adjusting the magnetic fields, requires a certain amount of time. Producing high resolution images, such as the ones in Figure 2.3, can take several minutes. Naturally, the quality of the images is strongly affected by inhomogeneities in the magnetic fields, the internal magnetic interactions and electromagnetic interference from the environment.

Since MRI is virtually noninvasive and is able to produce high quality images, it has quickly become very popular in structural imaging. These properties are crucial also in functional imaging, but the measures used in structural MRI, such as the proton density, are not directly related to neuronal activation. Coincidentally, the scanning parameters can be tuned so that the resulting images provide a measure related to the oxygenation level in the tissue. The measure is based on the differing magnetic properties of oxygenated and deoxygenated hemoglobin molecules, as further explained in Section 2.3.1. The idea in functional MRI is to record a sequence of such images at different time points to allow the local changes in oxygenation level to be analyzed.

Problems arise from the long duration of the scanning. Scanning intervals of several minutes would not allow accurate analysis of the activity in the brain. Additionally, movement of the head and other physiological changes during the long exposures would distort the images. Fortunately, the scanning parameters, mainly the timing of the radio pulses, can be adjusted to allow much faster scanning. However, the spatial resolution suffers greatly from such adjustments. For example, the relaxation process is not allowed to fully complete, resulting in much weaker signals. Therefore, the setup used in fMRI is a careful compromise between fast scanning and high resolution images. Current fMRI scanners are able to produce full head volumes with a time interval of a few seconds, but the spatial resolution is only a fraction of that used in structural imaging.

2.3.1 Measuring Hemodynamic Responses to Stimuli

The detection of changes due to neuronal activation in fMRI is based on the differing magnetic properties of oxygenated (*diamagnetic*) and deoxygenated (*paramagnetic*) hemoglobin molecules. Neuronal activation results in a localized change of blood flow and oxygenation levels, which can be measured using suitable scanning parameters. This produces a measure called blood oxygena-

tion level dependent (BOLD) signal (*c.f.*, Ogawa et al., 1992). These vascular or *hemodynamic* changes are related to the electrical activity of neurons in a complex and delayed way. Local changes in the level of oxygenation reveal the active areas, but it is not possible to completely recover the electrical processes from the vascular ones. The hemodynamic changes are hard to model, but as their nature is somewhat slow and smoothly varying, a Gaussian model is often used. Some newer models of the hemodynamic response function are actually based on measurements from the real brain.

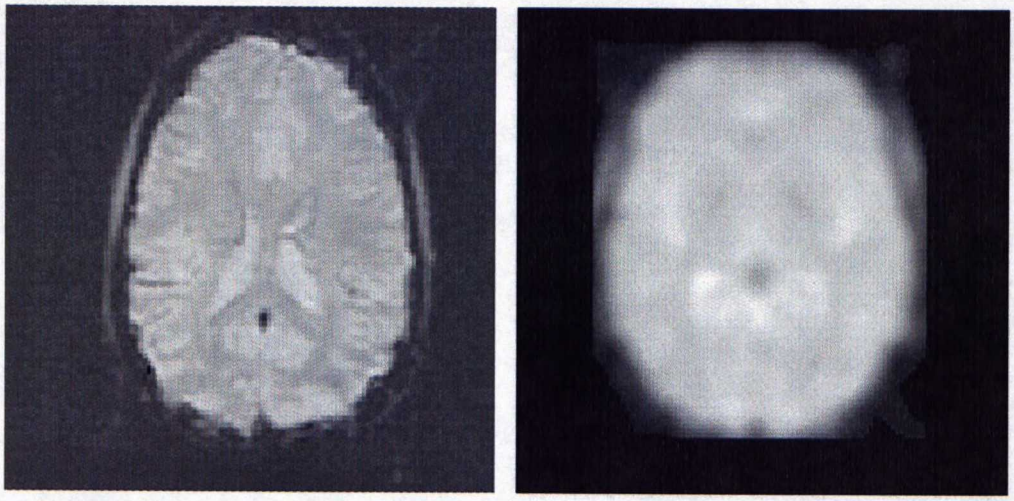
Usually, fMRI studies use a controlled stimulus, like visual patterns or audible beeps, designed to test a specific hypothesis. The simplest way of doing this is to repeat the stimulus several times, with resting periods in between, and scan the fMRI sequence during the whole time. Such experiments should reveal areas of the brain that are always active during the stimulation and inactive during the resting. Sometimes the subject can even be asked to perform a relatively simple mental or motor task during the scanning. Naturally, such tasks should not be allowed to result in head movement.

One example of an fMRI study is shown in Figure 2.4(a). The low resolution and the scanning parameters, optimized for BOLD, make the contrast between different tissue types very poor. Additionally, the fast scanning and low signal-to-noise ratio of the BOLD signal make the image very noisy. Therefore, a high resolution structural MRI is often scanned separately to aid in locating the activation during analysis by super-positioning. The bright areas in the images do not necessarily correspond to the active ones. Careful analysis of the whole sequence is required to detect the activation patterns.

2.3.2 Standard Preprocessing of Images

In addition to the low signal-to-noise ratio and additive noise, seen in Figure 2.4(a), the fMRI measurements are contaminated with artifacts, such as head movement and physiological vascular changes. Thus, the detection and analysis of interesting phenomena is very difficult. To overcome these difficulties, the images need to be preprocessed (*c.f.*, Worsley and Friston, 1995, SPM, 1999). Figure 2.4(b) shows an example slice after preprocessing. The level of noise is clearly reduced and the values are much more continuous. Also, the excess area outside the brain has been removed, and is shown in black. The usual steps of the preprocessing include:

- Retiming the slices to account for the fact that each slice was scanned at a slightly different time. Since the scanning is not instantaneous this kind



(a) (b)

Figure 2.4: Examples of functional magnetic resonance images. (a) A scanned slice without processing and (b) after the standard preprocessing has been applied. Note that the images do not show a direct measure of activation.

of temporal smoothing improves the continuity of the measurements.

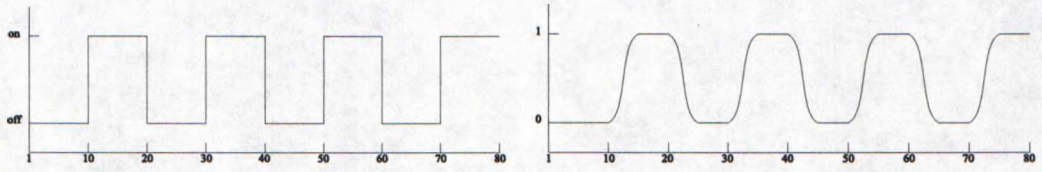
- Realigning the slices to reduce head movement related artifacts. This makes the voxels of different time points match spatially and removes some of the empty area surrounding the head in the images.
- Smoothing or low-pass filtering to reduce high frequency noise and to increase local voxel correlations. This increases the signal-to-noise ratio of the hemodynamic effect and makes detection of interesting phenomena easier.
- Normalizing the volumes into a standard coordinate system. This allows comparing different scans and identifying known locations.

The normalization step is not required for analysis, but is often done to match the functional volumes to the structural ones. Additionally, the normalization makes comparisons between different subjects easier. In practice, this may somewhat distort the brain in the images and individual differences may still remain quite big, so that caution needs to be taken when drawing inter-subject conclusions.

2.3.3 Standard Analysis of fMRI Sequences

The standard way of analyzing an fMRI sequence is to use statistical parametric mapping (SPM) (SPM, 1999), which is based on a general linear model (GLM) (*c.f.*, Worsley and Friston, 1995). Essentially, the analysis reveals the areas of the brain that most probably fit a given hypothesis, which is presented as a reference time-course.

The reference time-course can be approximated using the stimulation pattern and a model of the hemodynamic response. An example of such a reference time-course is shown in Figure 2.5. The depicted pattern is a very simple case of repeated on-off type of stimulus. The stimulation time-course is then convolved with the model of the hemodynamic response, assumed Gaussian in the illustration.



(a) (b)

Figure 2.5: Illustration of an ideal stimulus. (a) An ideal stimulation time-course comprising of rest and activation periods. (b) The corresponding ideal reference time-course convolved with the hemodynamic response. In reality the detected time-course varies significantly.

The analysis can be considered in two steps. First, the reference time-course is compared to the time-course of each voxel in the fMRI sequence statistically. This produces an image of the probability to fit the given time-course, where the voxels with the highest probabilities are considered to be active. However, the probability image is very noisy and the second step is to segment it into the inactive and active areas. The segmentation is made robust by using a statistical model for the noise, usually assumed Gaussian. The difficulty with this approach is to define a threshold for the probability of activation that produces an accurate segmentation. Choosing a too high value value easily leads to discontinuous or too small areas. On the other hand, a small value may produce big areas that do not accurately locate the activation of interest.

After the spatial activation patterns have been formed, the true activation time-course of each area is formed by taking the mean sequence of all the voxels in the area. Again, if the segmentation is poor, for example, due to an incorrect threshold value, the time-courses are not generated accurately.

There are big problems with such analysis. The accuracy is limited by the ability to approximate the parameters needed for the statistical fitting. Also, small changes to the parameters can change the results severely. Additionally, the stimulation setup has to be simple enough to allow predicting the responses, and forming the reference time-courses, in the first place. Therefore, detecting previously unknown phenomena is extremely hard or close to impossible. These are some of the reasons why current research focuses on more data-driven and adaptive methods, like independent component analysis, explained in Chapter 3.

2.4 Individual and Group Studies

The formulation of general hypotheses is possible only by comparing the results of many subjects. Functional brain studies, like all studies in the biomedical field, are often group studies where many patients are subjected to the same conditions or given the same treatment. However, even under a controlled environment the individual responses change, for example, with attention.

The inherent differences in both individual and group results make the studies difficult. Therefore, it is extremely important that the analysis methods themselves are consistent and reliable. One such consistent method is proposed in Chapter 4.

Chapter 3

Independent Component Analysis

3.1 Motivation

Imagine a room with many sound sources, for example, multiple people speaking simultaneously and perhaps some music in the background. The sounds in the room are recorded using multiple microphones. These recorded signals are weighted sums of the original signals emitted from the different sound sources. The mixing of the signals depends on, for example, the location of the microphones or the acoustic properties of the environment. It would be very useful to be able to estimate the original source signals from the observed mixed signals alone. As an illustration, consider the source signals in Figure 3.1(a) and the microphone recordings in Figure 3.1(b). These are not very realistic sound signals, but should illustrate the problem clearly. It may seem impossible to estimate the source signals from the observed ones.

The idea of solving the original source signals using only the observed signals with unknown mixing and minimal, if any, information on the sources is called blind source separation (BSS) (*c.f.*, Cardoso, 1990, Jutten and Herault, 1991). If the mixing were known, the problem could be solved using classical methods. Yet, the problem of solving both the mixing and the original sources at the same time is considerably difficult. Approaches to solving it need to make some, hopefully minimal, assumptions. For example, assuming that the sources contain significant autocorrelations allows the use of temporal decorrelation algorithms, such as SOBI (Belouchrani et al., 1993) or TDSEP (Ziehe and Müller, 1998).

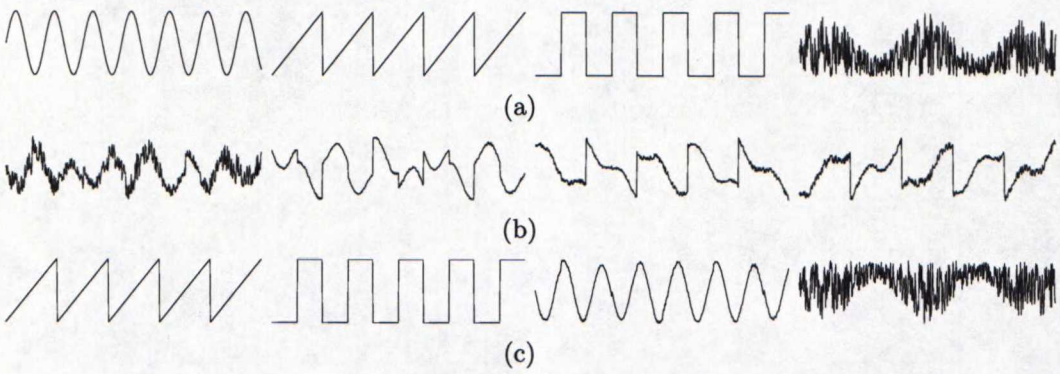


Figure 3.1: Illustration of source separation. (a) Each of the signals is emitted from a different source. (b) The recorded signals are differently mixed observations of the original source signals. (c) The source signals are estimated using only the observed mixed signals. The signals match very closely to the true source signals, with only minor differences like the reversed signs.

Independent component analysis (ICA) (*c.f.*, Jutten and Herault, 1991, Comon, 1994) is perhaps the most widely used method for performing blind source separation, and is implemented in many algorithms, such as FastICA (Hyvärinen, 1999) or Infomax (Bell and Sejnowski, 1995, Amari et al., 1995). It is based on the assumption that the source signals are statistically independent. This seems like a natural assumption in many applications and, in fact, it does not have to hold exactly in practice for ICA to work. Figure 3.1(c) shows the source signals estimated using FastICA, based only on the observed mixed signals shown in Figure 3.1(b). The estimated signals are very close to the original source signals. The minor differences, like the reversed signs, are explained in detail later. As a textbook for additional background information on independent component analysis and different algorithms consider reading Hyvärinen et al. (2001), Stone (2004).

3.2 Mixture Model

The model used in ICA is a statistical generative model for an instantaneous linear mixture of random variables. When the mixed signals are represented as a data matrix $\mathbf{X}$, the mixing model can be expressed in matrix form as:

$$\mathbf{X} = \mathbf{AS} \ . \tag{3.1}$$

Each row $\mathbf{s}_k^T$ of the source matrix $\mathbf{S}$ contains one independent component and each column $\mathbf{a}_k$ of the mixing matrix $\mathbf{A}$ holds the corresponding mixing weights, for a total of K sources. This is simply the matrix form of the linearly weighted sums $\mathbf{x}_t = a_{t1}\mathbf{s}_1 + \cdots + a_{tK}\mathbf{s}_K$.

Since both $\mathbf{A}$ and $\mathbf{S}$ are unknown, it immediately follows that the signs and scaling of the sources can not be identified. One can multiply the mixing vector $\mathbf{a}_k$ and divide the source vector $\mathbf{s}_k$ respectively with any coefficient. Additionally, the order of the sources is not fixed. Fortunately, the ambiguities in the model are often not so crucial. For example, the sign and scaling of the components are often normalized after ICA.

3.3 Estimating Independence

Theoretically, statistical independence means that the sources do not contain any information on each other. In other words, the joint probability density function (pdf) of the sources is factorisable on its marginal probability densities $p(\mathbf{s}_1, ..., \mathbf{s}_K) = \prod_i p(\mathbf{s}_i)$. Since a closed form solution to the ICA problem would require exact determination of the pdfs, which are generally not available, the sources have to be estimated by approximating independence with an objective function. This objective function can be based on concepts such as mutual information or negentropy (*c.f.*, Hyvärinen and Oja, 2000) and essentially measures how non-Gaussian the estimated sources are.

One way of defining the connection between independence and non-Gaussianity, as explained in Appendix A, is given by the common statistical measures of information theory. First, since mutual information measures the amount of information shared between random variables, it can be used as a natural measure of independence. Mutual information, in turn, is closely tied to negentropy, which basically compares a given density to a Gaussian. Finally, the difference to a Gaussian can also be approximated, without the exact pdfs, directly from the observations by using measures of non-Gaussianity, such as skewness and kurtosis, that is, the third and fourth order cumulants. For example, kurtosis is defined as:

$$kurt(s) = E\{s^4\} - 3E\{s^2\}^2 \ . \tag{3.2}$$

Another, more intuitive, explanation is offered by the central limit theorem. Basically, it states that the distribution of a mixture of *i.i.d.* random variables tends to be more Gaussian than the original ones. This means that, when the

sources are made more non-Gaussian, they must become more unmixed, or independent.

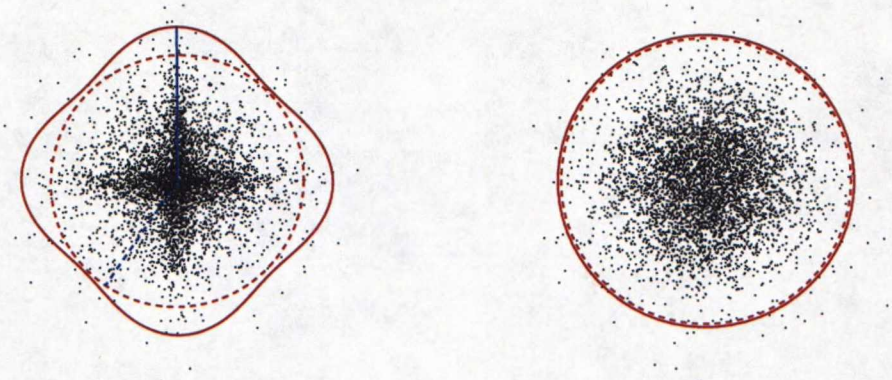
Before estimating the independent components, the observed data $\tilde{\mathbf{X}}$ can be whitened, that is, the samples made uncorrelated and their variances one. As explained in Appendix A, whitening is a linear transformation and can be constructed, for example, using principal component analysis (PCA) (*c.f.*, Jolliffe, 2002). The direction $\mathbf{v}$ of the first principal component is defined as the direction that maximizes the variance of the projection $\mathbf{v}^T\tilde{\mathbf{x}}$, where $\tilde{\mathbf{x}}$ is a column of $\tilde{\mathbf{X}}$. Generally, the ith principal component is along the direction of highest variance that is orthogonal to the previous $i-1$ components. One way of finding all the principal components is based on the eigenvalue decomposition (EVD) of the covariance matrix of the data, which gives two matrixes $\mathbf{D}$, a diagonal matrix of the eigenvalues, and $\mathbf{V}$, the corresponding eigenvectors.

The directions of the principal components define the uncorrelated basis and the corresponding variances give the scaling for the whitening transformation. Whitening the data does not constrain the ICA in any way, since the scaling is ambiguous anyway and independence implies uncorrelation. This is because uncorrelation means that the covariance is the identity, which always holds when the joint pdf is factorisable in the way shown before. The end result of the whitening is that the original mixing matrix $\tilde{\mathbf{A}}$ is also transformed in the following way:

$$\mathbf{A} = \mathbf{D}^{-1/2}\mathbf{V}^T\tilde{\mathbf{A}} \ . \tag{3.3}$$

Estimating the independent components from the whitened data matrix $\mathbf{X}$ is easier, since the number of free parameters is reduced. For example, the mixing matrix $\mathbf{A}$ is orthonormal, making its inverse $\mathbf{W}$ easy to calculate. If the whitening is done using PCA, the degrees of freedom can be lowered also in another way. By leaving out the weakest principal components the dimension of the data can be reduced in an optimal energy preserving manner, which improves the signal-to-noise ratio of the data.

Two examples of joint probability densities are shown in Figure 3.2. One is a mixture of arbitrary non-Gaussian densities, and the other one a mixture of Gaussians. The dashed curves around the densities plot the projected variance measured in all directions. The dashed line marks the direction of maximum variance, that is, the first principal component. Similarly, the values of kurtosis are shown using solid curves and the direction of maximum kurtosis with a solid line.



(a) (b)

Figure 3.2: Example joint propability densities. (a) For non-Gaussian densities the principal (dashed line) and independent (solid line) directions can be identified, where as (b) for Gaussian ones the directions are all equal. The corresponding dashed and solid curves show the values of variance and kurtosis in all directions respectively.

Looking at the arbitrary case, it is clear that ICA, based on the kurtosis, is able to identify the component directions much better than PCA. The principal components can also be identified and, as mentioned before, can be used in the whitening of the data. However, the directions of highest variance do not generally identify the independent components and the principal components are restricted, since they are always orthogonal.

On the other hand, the Gaussian case looks very different. Because a Gaussian density is perfectly symmetric and completely defined by the mean and standard deviation, the directions are all equal even with higher-order statistics, like kurtosis. Therefore, the results of PCA and ICA in such a case would be quite random. The special nature of the Gaussian density also means that, if more than one of the original sources is Gaussian, they can not be separated using ICA.

3.3.1 FastICA Algorithm

One particularly useful and widely used implementation of ICA is the FastICA algorithm (FastICA, 1998, Hyvärinen, 1999), which is very fast and robust. It performs well with big data sets and even under somewhat noisy conditions. FastICA uses a fixed-point optimization scheme based on Newton-iteration and an objective function related to negentropy. The remainder of this thesis fo-

cuses mainly on the use of FastICA, but the considerations should be relatively easy to extrapolate to other algorithmic implementations of ICA. FastICA can search for the independent components one at a time or all at once. Its performance can also be tuned somewhat by choosing from a range of nonlinearities. See Appendix A for more mathematical details of the fixed-point optimization.

The basic idea is to first whiten the data using PCA and then, based on the whitened data matrix $\mathbf{X}$, search for a solution in the form $\mathbf{s} = \mathbf{w}^T\mathbf{x}$, where $\mathbf{s}$ and $\mathbf{x}$ are columns of the source matrix and whitened data matrix, respectively. Or equivalently in matrix form:

$$\mathbf{S} = \mathbf{W}\mathbf{X} \ , \tag{3.4}$$

where $\mathbf{W} = \mathbf{A}^T$ is the demixing matrix. The algorithm optimizes the objective function, which estimates the sources $\mathbf{S}$ by approximating statistical independence. The algorithm starts from an initial condition, for example, random demixing weights $\mathbf{w}$. Then, on each iteration step, the weights $\mathbf{w}$ are first updated, so that the corresponding sources become more independent, and then normalized, so that $\mathbf{W}$ stays orthonormal. The iteration is continued until the weights converge. For example, when using the *cubic* nonlinearity, which corresponds to estimating kurtosis, the fixed-point update rule becomes (*c.f*, Hyvärinen and Oja, 1997):

$$\mathbf{w}^+ = E\{\mathbf{x}(\mathbf{w}^T\mathbf{x})^3\} - 3||\mathbf{w}||^2\mathbf{w} \tag{3.5}$$

3.3.2 Estimation Errors

It is clear that with such estimation schemes the solutions are only approximate and noisy. The true solution may not always be found. For example, in addition to the ambiguities in the ICA model, the strict assumption of statistical independence of the sources may not hold for a given data. Still, the model in Equation 3.1 is noise free, and noise often makes finding the optimal solution harder. Additionally, a problem is that the solutions can be affected by the parameters of the algorithm, for example, the initial conditions. Even the convergence stopping criterion may prevent the algorithm from finding the optimal solution. Moreover, algorithms often use clever adaptive tricks during iteration to be faster and more robust.

The estimation errors can be considered as an error-surface. The shape of the

surface is determined by the optimized objective function and the given data. For example, a 3-dimensional error-surface is shown in Figure 3.3, where in addition to the minimum point, the surface contains local minima and some noise. Different starting points and directions, that is, the initial conditions, cause the optimization to converge along different paths, shown as thick curves. Although a robust algorithm should not be affected by the noise, it can get stuck on a local minimum and even the optimal solutions reached along different paths can be slightly different.

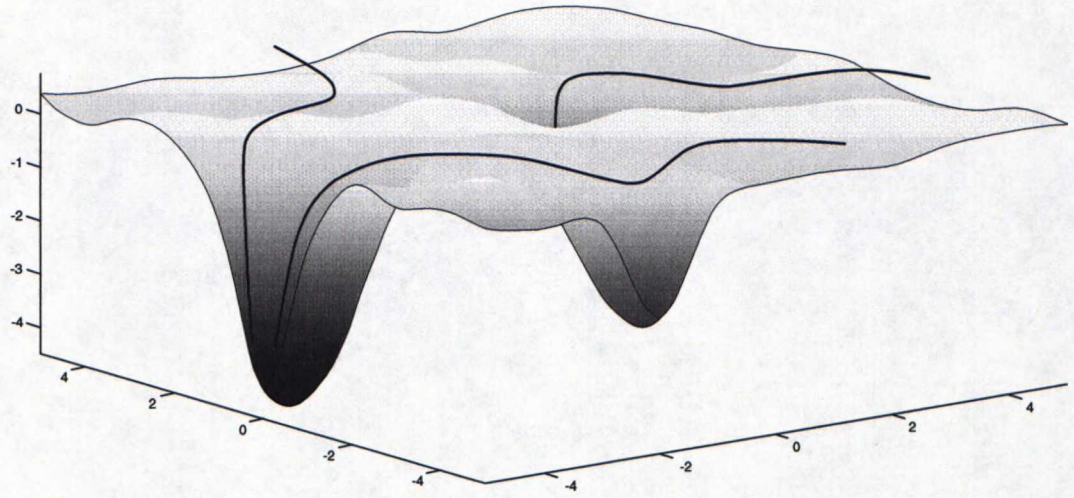


Figure 3.3: Error-surface of estimation. Different initial conditions cause the algorithm to converge along different paths (shown as curves) on the 3-dimensional error-surface. Thus, the algorithm may get stuck on a local minimum and even the optimal estimates may be slightly different.

Furthermore, since the number of free parameters can be very high, it is relatively easy for the ICA estimation to overfit the data. Overfitting can cause severe estimation errors. Using PCA to reduce the dimension of the data during whitening has been shown (Särelä and Vigário, 2003) to help in preventing the possibility of overfitting.

3.4 Application to fMRI Data

ICA was first applied to fMRI data by McKeown et al. (1998) and has since been used in many studies (*c.f.*, Jung et al., 2001, Kiviniemi et al., 2003). It has opened new possibilities in designing studies and analyzing measurements in functional brain research. To justify the application of ICA to the volumetric

(3-dimensional) spatial fMRI signal, the data must match the assumptions and limitations of ICA, at least sufficiently. This has been shown (McKeown and Sejnowski, 1998) to be the case. For more information and great overviews on the use of ICA in fMRI studies consider reading, for example, Calhoun et al. (2003), McKeown et al. (2003).

3.4.1 Spatial ICA

To be able to use the fMRI signal in the ICA model, each scanned volume must be transformed into vector form in a bidirectional manner. Since ICA is only concerned with the statistics of the observations and not the order of samples within them, the voxels in the volumes can be reordered into vectors quite freely (*c.f.*, Calhoun et al., 2003). Naturally, all volumes must be transformed using the same reordering.

Usually, ICA is applied to temporal signals, such as EEG or MEG recordings, in the same way as with the illustrative sound signals in Section 3.1. But, the fMRI signal is a temporal sequence of the scanned spatial volumes, and the different activation patterns are also spatial. Therefore, a transposed version of the ICA mixing model is used. The model is illustrated in Figure 3.4.

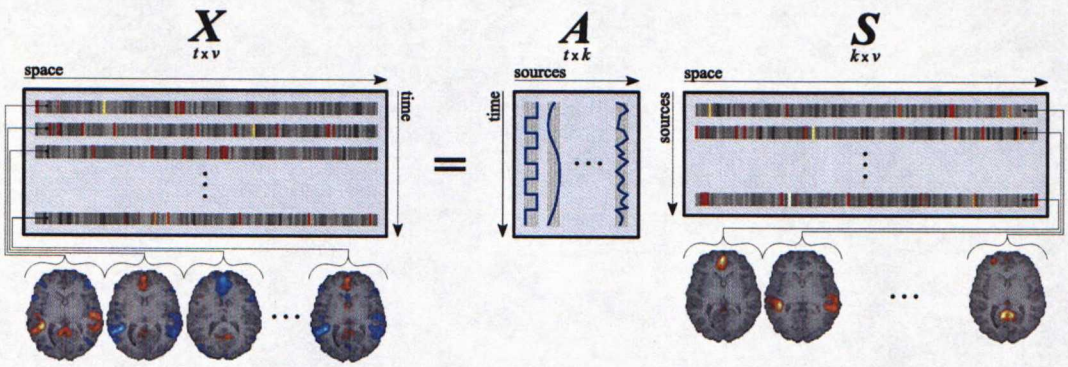


Figure 3.4: Spatial ICA of fMRI data. The rows of the data matrix $\mathbf{X}$ and sources matrix $\mathbf{S}$ are vectorized volumes. The corresponding columns of the mixing matrix $\mathbf{A}$ are the time-courses. Note, that the statistical independence applies to the volumes.

The fMRI signal is represented by a $T \times V$ data matrix $\mathbf{X}$, where T is the number of time points and V is the number of voxels in the volumes. This means that each row of $\mathbf{S}$ contains an independent spatial pattern and the corresponding column of $\mathbf{A}$ holds its activation time-course. Since the whole

mixing model is transposed, the statistical constraint applies to the spatial domain.

It is also important to note that, contrary to the traditional fMRI analysis method, the time-courses are not imposed in any way. In ICA they manifest themselves in a purely data-driven way. For more general information on the differences between supervised and unsupervised methods consider reading Haykin (1998).

Chapter 4

Variability of Independent Components

4.1 Motivation and Sources of Variability

Some analysis methods, for example, based on Bayesian inference, can provide information on the reliability of the solutions. ICA algorithms, however, do not readily offer any information on the reliability or stability of the solutions. Thus, understanding the causes of the variability and how to exploit it are essential for consistent analysis with ICA.

The sources of variability can be roughly categorized into three groups. First, theoretical assumptions behind the ICA algorithm may not fully hold (*c.f.*, McKeown and Sejnowski, 1998). For example, the assumptions of strict statistical independence or absence of noise may limit the ability to converge to the true optimum, as explained earlier in Section 3.3.2. Second, the algorithm itself may be a source of variability as a result of the used optimization scheme. For example, initial conditions or adaptive iteration may affect the accuracy of estimating the independent components. Third, as the data partly defines the error-surface, it also affects the variability. The analysis of consistency may, therefore, reveal more about the properties of the data.

Additionally, the whitening, and possibly reducing the dimension, of the data is important. If the degrees of freedom is made too small, there may not be enough information to fully characterize the variability. On the other hand, if it is too high, the ICA algorithm is likely to overfit the data and cause severe variability, since on each application of ICA, the overfits can change freely.

This Chapter presents a method that exploits the inherent variability of the estimated independent components to acquire additional insight on the data. It is based on running ICA multiple times, in a controlled way, and then further analyzing the estimated components. Additionally, the results of the analysis are visualized to allow fast and easy interpretation, as explained Chapter 5.

4.2 Exploiting Variability

To gather as much information on the variability as possible, ICA needs to be run multiple times. Naturally, a fast and efficient algorithm is required for keeping the analysis time meaningful. For example, the FastICA algorithm (FastICA, 1998) easily allows tens or even hundreds of runs to be used. The key is to adequately span the error-surface by randomizing the initial conditions and resampling the data on each run to allow the ICA algorithm to converge to different solutions. This should allow the identification of the consistent estimates and further analysis of the remaining variability.

4.2.1 Randomizing the Initial Conditions

Consider again the error-surface depicted in Figure 3.3. Randomizing the initial conditions makes the starting point and direction of optimization change on each run. This allows the ICA algorithm to consistently converge to the optimum solution by approaching it from different directions. Actually, some of the directions may allow a more optimal convergence than others and optimal points surrounded by local minima or noise may only be reached from certain directions.

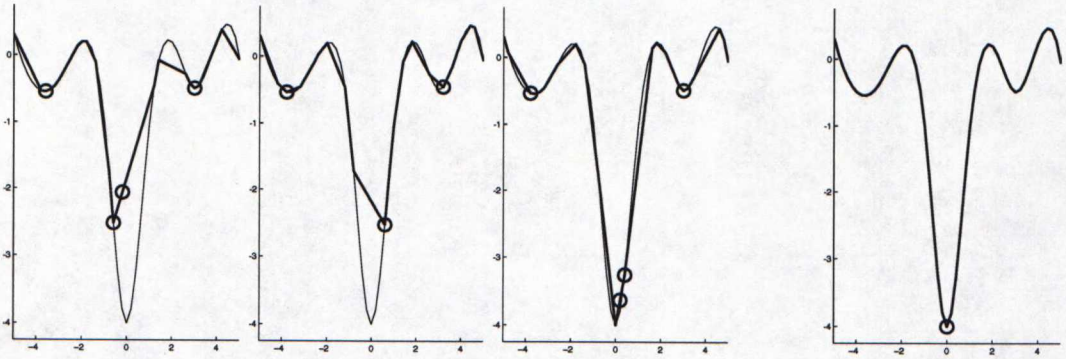
Thus, by randomizing the initial conditions the method is able to exploit the variability to improve estimates and detect components, which are difficult to find. Also, it should allow to work around the limitations of the ICA algorithm to some extent, for example, under a high level of noise.

4.2.2 Resampling the Data

Resampling the data is often called bootstrapping (Meinecke et al., 2002) and the idea is, quite simply, to randomly resample the data with reposition on each pick. The point is that the resampling should not affect the global statistical properties of the data, but the error-surface, partly defined by the

resampled data, would be slightly different on each run. The strong optimum point stays relatively stable, but the shape of local minima and noise on the surface can change freely. Thus, it allows the ICA algorithm to converge to different solutions and even reach solutions otherwise difficult to identify due to surrounding local minima or a high level of noise.

The idea is further illustrated in Figure 4.1 using a 2-dimensional error-surface, shown as thin curves. The data is resampled on each run, causing the shape of the surface to change slightly, shown as thick curves. This allows the algorithm to converge to different individual solutions, marked as "o". Some of the solutions can be around local minima, but after running the algorithm many times most of the solutions are near the true optimum. Therefore, the mean of all solutions gives the best estimate of the true optimum.



(a) (b) (c) (d)

Figure 4.1: The idea of bootstrapping. On different runs (a)–(c) the data is resampled (shown as thick curves) to slightly change the shape of the 2-dimensional error-surface (shown as thin curves). This allows the algorithm to converge to different solutions (marked as "o"). They may be around local minima, but after many runs (d) the best solution is given by the mean of the different runs.

4.3 Analyzing Consistency

The solutions of multiple runs are reached along different paths on the error-surface, based on the initial conditions and the resampled data. In theory, when the same solution is reached in more than one run, the estimates would match perfectly. This is often not the case, but the estimates should still form consistent groups with high similarity. Thus, one may expect that the true solution can be found as a mean of the consistent group and that the spread of

the group can be used to analyze the reliability of that solution. On the other hand, if the spread is too high, one expects the solution not to be reliable. The mean representatives of the groups can also depart somewhat from the strict assumption of independence, which may actually lead to a more natural decomposition of the data.

4.3.1 Multiple Runs of ICA

Early proposed bootstrapping of ICA use a standard initial run. Then, consecutive runs are made by resampling the estimated independent components from the initial run. The reason for this is that the solutions of the consecutive runs become only slightly altered versions of the already independent solution and comparing the solutions is easier, due to mixings being close to unity and permutations easier to handle. The biggest drawback, however, is that the initial run defines the set of independent components to be analyzed. Also, the variability is more restricted so the error-surface may not be spanned sufficiently and a poor solution on the initial run would render the consecutive runs useless.

On the other hand, when the resampling is done without an initial run (*c.f.*, Duann et al., 2003) the total number of runs is often very small and comparing the solutions is done by hand. Furthermore, randomizing the initial conditions is sometimes neglected in bootstrapping approaches since some ICA implementations are not very sensitive to them.

The presented method does not use an initial run, allowing any permutation, and indeed, any combination of independent components to be found on each run. The benefits are that ICA has more freedom to find the optimal solution on each run and the solutions can include components that are very difficult to find with just a single run or other bootstrapping methods.

After multiple runs, the easiest way to describe the results of ICA, and maybe the only way due to memory constraints, is to use a concatenated version of the mixing matrixes:

$$\mathbf{A}^{all} = [\mathbf{A}^1|\mathbf{A}^2|\cdots|\mathbf{A}^B] \ , \tag{4.1}$$

where $\mathbf{A}^b$ is the mixing matrix from run b and B is the total number of runs. Thus, each column of the matrix $\mathbf{A}^{all}$ defines one estimated independent component.

Additionally, the size of the data matrix can be reduced by taking only a given

percentage of samples during the resampling. This makes ICA considerably faster, allowing a clear increase in the number of runs used to analyze the consistency. The reduction of samples may eventually increase the variability somewhat, but with a properly selected amount of samples the overall quality of the method should increase due to the multiple runs. On the other hand, the reduction of samples also lowers the possibility of overfitting.

The reduction of samples is justified by the experiment shown in Figure 4.2. It shows how the variability is increased when the amount of samples is lowered. The amount of variability is measured as the mean difference of independent components between estimates using only a percentage of the samples and estimates using all of the samples. This is accomplished by running ICA 50 times with each percentage value and calculating a set of mean component estimates for each percentage. Then, the difference of the mean sets compared to the set using all of the samples is calculated using the Frobenius norm.

The horizontal axis shows the amount of samples used and the curves show the cumulative differences of 5 independent components, that is, the bottom curve is the mean difference of the first component, the second curve the cumulative difference of the two first components, and so on. Clearly, the amount of variability is very stable at higher percentages and starts to increase only at very low values. Therefore, the fraction of samples used can be quite small, for example, 20 percent, without significantly increasing the amount of variability.

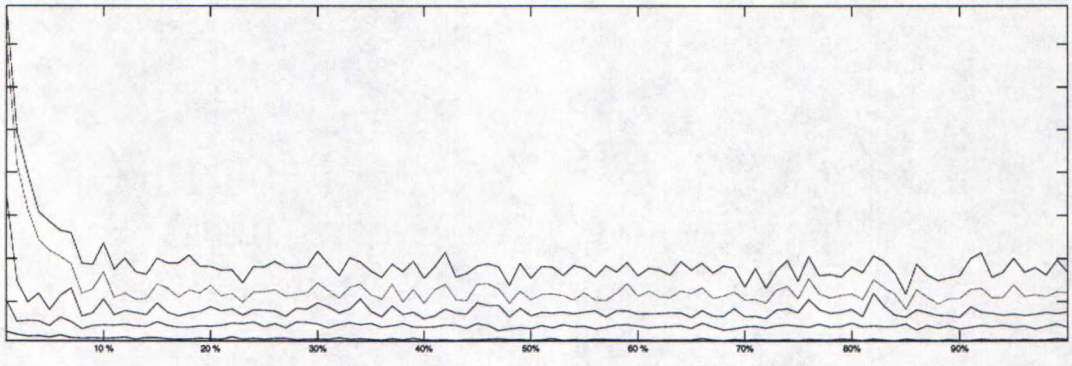


Figure 4.2: Controlling variability by resampling. The cumulative mean differences (shown as curves) of 5 independent components measured between estimates using the shown percentage of samples and estimates using all of the samples. The bottom curve is the difference of the first component. The second curve is the cumulative difference of the two first components, and so on.

4.3.2 Clustering the Estimates

Before the estimates from Equation 4.1 can be clustered and the consistency of the solutions analyzed, the similarity of the estimates must be measured. First the estimates are normalized to account for the scaling ambiguities of ICA. One good way of doing this is to make the estimates have zero mean and unit variance. If the normalized solutions are defined as $\hat{\mathbf{A}}$, with columns $\hat{\mathbf{a}}_i$, the normalization equals:

$$\hat{\mathbf{a}}_i = \frac{\mathbf{a}_i - \bar{\mathbf{a}}_i}{||\mathbf{a}_i - \bar{\mathbf{a}}_i||} , \tag{4.2}$$

where $\bar{\mathbf{a}}_i$ is the mean of $\mathbf{a}_i$ and $\mathbf{a}_i$ is a column of $\mathbf{A}^{all}$ respectively. Now, the only remaining ambiguities in $\hat{\mathbf{A}}$ are the permutation and sign. The best way to measure the similarity of these normalized results is through correlation:

$$\mathbf{C} \propto \hat{\mathbf{A}}^T \hat{\mathbf{A}} . \tag{4.3}$$

Then, since only significant similarities are considered interesting, the correlation matrix $\mathbf{C}$ can be thresholded with a suitable threshold value ϵ as:

$$\tilde{c}_{ij} = \begin{cases} 1, & \text{if } |c_{ij}| > \epsilon \\ 0, & \text{if } |c_{ij}| \leq \epsilon \end{cases} . \tag{4.4}$$

This leads to a binary correlation matrix $\tilde{\mathbf{C}}$, whose elements $\tilde{c}_{ij}$ mark whether the two corresponding components can be considered as estimates of a common underlying source.

Furthermore, if the binary correlation matrix $\tilde{\mathbf{C}}$ shows that estimate $\hat{\mathbf{a}}_i$ is related to $\hat{\mathbf{a}}_j$, and that $\hat{\mathbf{a}}_j$ is to $\hat{\mathbf{a}}_k$, this is a strong suggestion that $\hat{\mathbf{a}}_i$ is also related to $\hat{\mathbf{a}}_k$ through a longer path, even if $\hat{\mathbf{a}}_i$ and $\hat{\mathbf{a}}_k$ exhibit a somewhat lower correlation. This allows somewhat weaker and discontinuous paths also to be taken as estimates of a common underlying source. To archive this, the binary correlation matrix can be raised to a suitable power, whose value p relates to the length of the longest acceptable path:

$$\mathbf{R} = \tilde{\mathbf{C}}^p . \tag{4.5}$$

See Figure 4.3 for illustration of the different phases of the correlation calculations. It shows the succession from absolute correlation values $|\mathbf{C}|$ through

thresholded binary correlations $\tilde{\mathbf{C}}$ to relations $\mathbf{R}$ after raising to a power. Finally, the relations are shown in an order that makes the groups more compact and easier to evaluate. The values are shown as grayscale images with white being 0 and black 1. In particular, note how raising the binary correlations to a power makes the subsets in the data more visible.

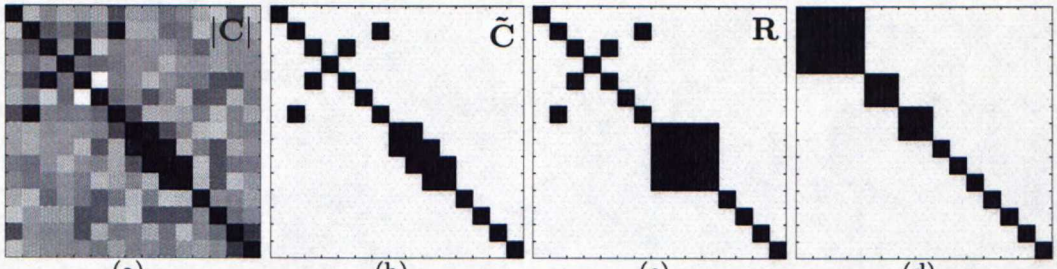


Figure 4.3: Phases of correlation calculations. Grayscale images with white being 0 and black 1 of the succession from (a) absolute correlation values $|\mathbf{C}|$ through (b) thresholded binary correlations $\tilde{\mathbf{C}}$ to (c) relations $\mathbf{R}$ after raising to a power. (d) The final relations sorted so that groups are more compact.

When the estimated components $\hat{\mathbf{A}}$, their correlations $\mathbf{C}$, and their binary relations $\mathbf{R}$ are available, the components can be clustered into consistent groups. However, unlike other methods (*c.f.*, Icasso, 2003, Himberg et al., 2004) this is not done using a general clustering method, such as hierarchical clustering or k-means.

The problem with general clustering methods is that they are very time consuming and, in a way, overcomplete. The complete hierarchy, often shown as a dendrogram, may actually confuse more than help interpret the results. General methods may not perform well when using correlation as the distance measure. They are also often too discriminative against clusters, which do not form compact groups, such as subspaces or the longer paths of correlation. Additionally, they may require some parameters that are unknown and would need to be estimated, like the correct number of clusters needed in k-means.

Thus, the clustering is done with a method that is simpler and more suitable for the situation. It does not need any additional parameters and is based on the fact that each element c_{ij} of matrix $\mathbf{C}$, and r_{ij} of $\mathbf{R}$ respectively, links two estimated components $\hat{\mathbf{a}}_i$ and $\hat{\mathbf{a}}_j$. The method works in two phases:

1. All elements that are related, that is, have a value r_{ij} of 1, are sorted into a list in descending order of the corresponding absolute correlation value

$|c_{ij}|$. Thus, the resulting list will hold the links between all component pairs in their order of importance.

2. Then, the clusters are formed by going through this list in order until all components have been clustered by making the following decisions for each link:

   - If neither of the linked components $\hat{\mathbf{a}}_i$ and $\hat{\mathbf{a}}_j$ belongs to an existing cluster, create a new cluster containing both components.
   - If one of the linked components $\hat{\mathbf{a}}_i$ or $\hat{\mathbf{a}}_j$ belongs to an existing cluster, add the other component to that cluster as well.
   - If both of the linked components $\hat{\mathbf{a}}_i$ and $\hat{\mathbf{a}}_j$ belong to an existing cluster, do nothing.

4.3.3 Properties of the Groups

As explained before, when the estimated components have been grouped, the consistent components can be identified using the variability within the groups. The mean of each group can be calculated as the best estimate of the true decomposition. The spread of the grouped estimates around the mean reveals the magnitude and shape of the variability. Moreover, additional properties of the groups can be calculated to help human interpretation. For example, measures of compactness and discrimination of the groups can be very helpful, since they can provide clues to the nature of the variability. Some of the properties can be used to rank the components and sort them accordingly. With a good ranking, the human interpretation becomes faster, since the most interesting components appear first, and similar phenomena can be shown close to each other.

Naturally, the difficulty is how to define a ranking that favors the interesting components. In discriminant analysis, the ratio between the distance to other groups and the size of a group, in the Euclidean sense, is often used. However, the number of estimates in each group can be very different, offering evidence of the relevance of the group. Therefore, such a rank should be weighted with the number of estimates in each group. For example, when two groups are equally well discriminated, the one containing more estimates should be considered more consistent. Additionally, as the rank is a simple value, it should be made more robust to outliers by using the geometric mean when calculating the Euclidean distance ratios. One way of ranking that performs

sufficiently well, can be defined, for group i as:

$$\beta_i = \log(1 + c_i \frac{d_i^\circ}{d_i^\bullet}) \ , \tag{4.6}$$

where c_i is the proportion of estimates in the group, $d_i^\bullet$ is the normalized geometric mean of the intra-group distances and d_i° is the normalized geometric mean of the inter-group distances. The normalizations are needed to make the different measures comparable and the logarithm simply increases the contrast of the ranking.

The most reliable, that is, the most consistent and well discriminated, components are often the most interesting. However, there may be some exceptions, for example, components that are very difficult to find. Being able to more truly define what makes a component interesting would allow a very useful ranking indeed. Recently, in the field of biomedical signal processing, some quantitative methods to analyze the nature of signals have been developed. Some of them (*c.f.*, Gautama et al., 2004) try to classify the nature of a signal based on how linear the signal is or whether it is deterministic or stochastic. Others (*c.f.*, Formisano et al., 2002) use entropy to measure how structured and, therefore, interesting the signal is. It may be possible to construct a ranking by using such methods also on the temporal activation patterns of the components.

4.4 Interpreting the Variability

After the estimates have been grouped, and the additional properties calculated, the components can be interpreted reliably. Since the information is multi-modal and its amount can be quite overwhelming, human interpretation requires it to be shown in a clear and easy to understand form. This is, by all means, not trivial and the visualization is explained separately in Chapter 5.

As shown in previous work (Ylipaavalniemi and Vigário, 2004), the nature of the variability can be very different in each component and provides important additional information. The spread of a group around the mean time-course can be almost constant at all time points or it be very structured. For example, it can follow the time-course very closely or be related to the time-course of another component or the stimulation pattern. Additionally, there can be single time points with very high variability, even when the main portion would be consistent. The nature of the variability, combined with the measures of

discrimination, can be used to easily detect artifacts or characterize other phenomena. Many examples are shown in Chapter 7.

Considering the whole independent volumes can also be beneficial, although more time consuming, since similar interpretations can be based on the spatial variability. In particular, localized features may be easier to interpret in the spatial sense. Moreover, the regions with strong variability could be masked out of the data relatively easily by leaving those voxels out of the analysis. This should improve the estimation of the independent components based of the remaining data.

Instead of completely discarding, as in masking out, the regions with high variability, those areas could be analyzed using different methods. For example, methods not based on statistical independence may be able to reveal properties that ICA is not able to decompose. The recently introduced framework of denoising source separation (DSS) (Särelä and Valpola, 2005) uses denoising functions to identify the signal decomposition. Such functions can be defined as weighting masks and the spatial maps of strong variability could be used to build them. Additionally, the DSS estimation does not have to be based strictly on statistical independence.

Chapter 5

Visualization

5.1 Relevance

Since the original data and decomposition are three-dimensional volumes, showing only a time-course or a planar slice of a volume would not give an accurate view of the data or results. Therefore, it is important to visualize the information in a form that allows all the information to be analyzed easily. This is not simple since the amount of data is huge, easily making the visualization cluttered or too slow to be of any value.

Fortunately, interpreting volumetric data is quite natural for humans and specifically in medical sciences, doctors are accustomed to using images, such as ordinary x-rays. The decomposition also gives multi-modal information consisting of the volumetric functional patterns and the related activation time-courses. Visualizing the information in a clear and natural way is very important for fast and easy interpretation.

Due to their efficiency and possible interest to many other analysis environments, the visualization tools created during the work are considered as a possible future toolbox to be published on its own. Open questions related to the release of such a toolbox concern mainly portability to other environments and interoperability with existing tools.

5.2 Showing Brain Activation

The whole point of the analysis is to locate activated areas in the brain. Quite naturally, the aim of the visualization must be to allow the user to accurately see these locations. This can be problematic, as the activation patterns can be of any shape and size. Additionally, due to their smoothing preprocessing, the structural information in the functional volumes is very poor. Therefore, it is best to overlay the functional patterns on top of a structural MRI, if available. This allows the user to interactively study the content of the volumes, comparing the brain activation to known regions in the structural image. The interactive nature of the visualization adds constraints to the system in terms of usability. For example, updating the images on the screen has to be fast enough.

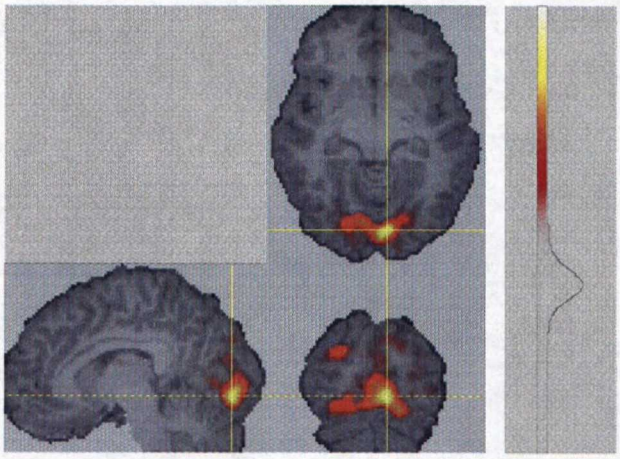
The estimated independent components have a time-course in addition to the spatial pattern. The temporal information is as important in the analysis of fMRI data as the spatial, because the time-courses reveal how the activation patterns are related to the stimuli. Additionally, the information acquired from the multiple runs, related to the grouping of the estimates and their nature, can be very helpful during the interpretation. For example, it is easier to focus on the most relevant results by spotting the most consistent, or reliable, components.

5.2.1 Spatial Information

Part of the interactive visualization interface is shown in Figure 5.1(a), with an example of activation pattern under study. The interface shows the volumes simultaneously from three orthogonal directions, aligned with the main axes of the volume. The three slices are linked together and the yellow cross-hairs pinpoint the current location, which lies in the intersection of the three slices. The user can move the current location freely to any point in the volume. A structural MRI is shown as a template in grayscale and the functional volume is overlaid on top of it in color.

The coloring is based on a smooth gradient, that is, a lookup table for a range of smoothly interpolated values, which makes stronger activation show up in brighter (hotter) colors. The color gradient can be seen in Figure 5.1(b), which shows the histogram of the activation volume. As the active regions are sparsely distributed, most of the volume is, in fact, noise. Therefore, the main lobe of the histogram can be considered as noise, or the inactive region.

The volume is always shown so that the tail of the histogram with the most energy, or mass, is considered to be the positive extreme. This effectively fixes the sign ambiguity of ICA. However, there are cases where the histogram is almost symmetric, or nearly all of the energy is in the main lobe. This may suggest that, in such cases, there is no significant focal activation in the volume. For example, some artifacts seem to produce such a volume.



(a) (b)

Figure 5.1: First partial view of the interactive user interface. (a) The volumes from three orthogonal directions with the cross-hairs pinpointing the current location. (b) The histogram of the activation volume. The structural volume is drawn using grayscale values and the amount of activation with a color gradient, where brighter (hotter) colors mean stronger activation.

The gradient used for coloring the activation pattern is fitted to the range from the main lobe to the more powerful extreme. The lower end of the gradient is fully transparent and the higher end fully opaque. As mentioned, the color also changes smoothly from darker (colder) to brighter (hotter) values. For example, it is very easy to see that a strongly activated area is located at the back of the brain on the right side, most probably related to processing visual information.

Additionally, the whole spatial patterns produced by ICA are, by definition, independent of each other and thresholding the activation patterns is not as crucial as in the traditional fMRI analysis method, explained in Section 2.3.3. When the thresholding does not essentially change the results, it is easier to overlay the activation pattern on the structural image using a smooth level of transparency. This makes it easy to see how noisy the volume is or if there are many separate areas of activation. For example, often the two hemispheres of

the brain are activated symmetrically, but the dominant side contains stronger activation.

5.2.2 Temporal Information

Interpreting the true nature of the variability from numerical values, like variance and rank, alone is practically impossible. Visualizing the distribution of the grouped components around the representative mean time-course of the group, the magnitude and nature of variability in the group become immediately clear. Figure 5.2 shows an example time-course of a component with the distribution drawn around it. The different quantiles are drawn using different grayscale values. For example, it is very easy to see that this component is quite consistent. The spread is big only at some time points, near sharp transitions, and only the extreme quantiles, that is, lightest shades of gray, seem to have that bigger spread.

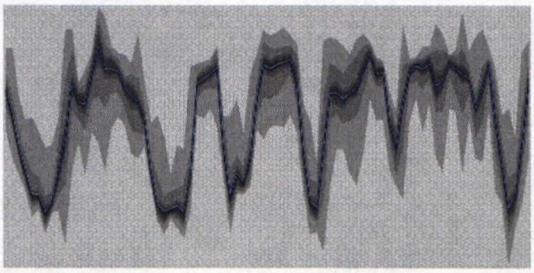


Figure 5.2: Second partial view of the interface, showing the activation time-course and its variability. The distribution of the group is drawn around the mean time-course using grayscale values for quantiles.

Naturally, it is also possible to calculate and show the variability from the volume, but calculating the volumetric distribution is very time consuming and may not always be worth the effort. Still, the volume may allow easier interpretation of certain kind of variability, as shown in one of the results from the experiments in Figure 7.5, on page 51.

5.2.3 Group Information

The components are ranked using a complex measure that produces a meaningful ordering, as explained in Section 4.3.3, but it can be very difficult to understand how the groups really differ. The problem is that the rank is always

a compromise between the parameters in the measure. However, in addition to the numerical values that tell how well a group is discriminated, that is, separated from the other groups, the measures can also be visualized for easy interpretation. An example of such visualization is shown in Figure 5.3. The disks show the spread of Euclidean distances with the minimum distance defining the inner radius and the maximum the outer radius. The circles over the disks mark the mean distance. The left disk shows the spread of intra-group distances, measured between the members of the group, and the right disk shows the spread of inter-group distances, or the distances to all other groups.

This type of information is very easy to interpret. For example, the relatively small intra-group disk tells that the group is quite compact, or consistent. Specifically, the mean value is very small, suggesting that there are only a few outliers with higher distances. The inter-group distances also form a rather compact range around a large mean value, meaning that the group is well separated from all the other groups. Additionally, the good discrimination of the group can be seen from the fact that the left disk would fit completely inside the hole in the right disk, with some room to spare.

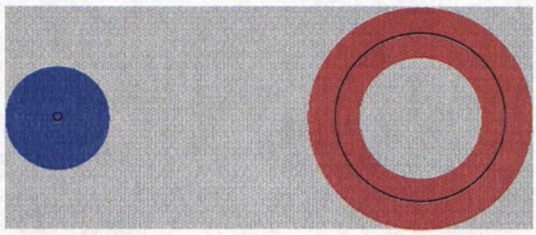


Figure 5.3: Third partial view of the interface, showing the discrimination power of the group. The left disk shows the spread of the Euclidean distances between the members of the group and the right disk between the distances to all other groups, with the black circles marking the mean values of those distances.

5.3 Complete User Interface

All the previously discussed multi-modal information is brought together to allow the interactive human interpretation. The complete user interface, for a single component, is shown in Figure 5.4, which combines the parts shown in Figures 5.1, 5.2 and 5.3. The interface includes some numerical properties of the component, shown above and below the time-course. Also, a possible reference time-course is shown as a two color pattern beneath the activation

time-course. The bands depict the on-off nature of the stimulus. This enables the user to better see how the component is related to the stimulus.

The numerical information is related to the grouping and the histogram of the activation pattern. The number of estimates in the group and the normalized rank of the component are on the top. The ratio of energy between the upper and lower tail of the histogram, related to skewness, and the amount of energy in the main lobe of the histogram are on the bottom. More examples of the visualization, and related interpretations, are shown in Chapter 7.

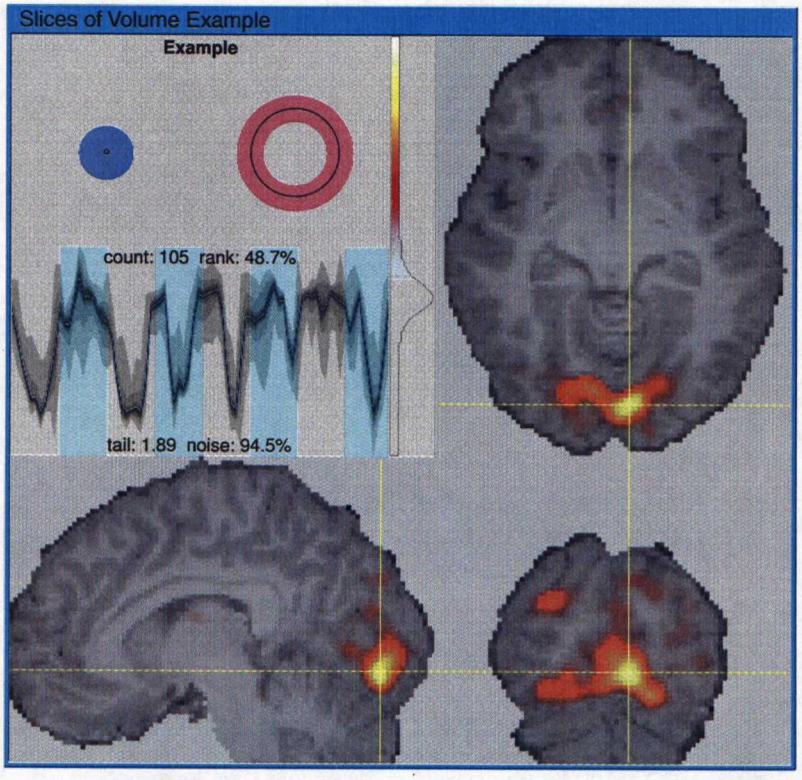


Figure 5.4: The complete view of the interactive user interface, showing all the information related to a single component.

Additionally, the user interface offers helpful interactive tools, which make it very user friendly. For example, the functional overlay can be toggled on or off to reveal the structure underneath. The activation pattern can also be viewed without the structural template, and without the artificial coloring, which can be very useful in situations where the component does not contain clear, or focal, activation. Also, the current location can be centered on the strongest activation automatically. All the interactive tools are focused on making the interpretation fast and easy.

5.4 Medical Standards

The medical field has a lot of information on the brain from the increasing number of studies conducted. Some of this information is available in brain atlases, which use a standard convention to localize areas of the brain, such as Talairach coordinates (Talairach and Tournoux, 1988) or the older Broadman's areas. Incorporating such information into the visualization would further increase the usability of the tools and reliability of the interpretations.

Chapter 6

Experimental Setup

6.1 Real fMRI Data

To test the usefulness of the ICA analysis and visualization method, explained in Chapter 4, the data from a real functional magnetic resonance imaging (fMRI) study was analyzed. It is important that the method is tested with real world data, since an artificial data set would not prove the usefulness under realistic noise and variability structures. Additionally, the method might reveal unseen phenomena from the data.

The experiments involved 14 voluntary subjects, which are referred to only by their initials to protect their anonymity. The stimulus and scanning conditions were repeated as closely as possible for each subject to allow the comparison of results between subjects.

6.1.1 Auditory Stimulation

The study used an auditory stimulus, which consisted of repetitions of spoken text with resting periods in between. These periods were repeated four times during the experiment. The results should reflect this by showing activation at least on the primary auditory areas, but also on additional language and memory related areas. The activation time-courses of these areas should be somewhat related to the stimulus, especially on the primary areas.

6.1.2 Volume Acquisition

During the four repetitions of speech and resting periods 10 full head volumes were acquired in each condition with a scanning interval of approximately 3 seconds, resulting in a total of 80 volumes. The scanning was done using a 3 Tesla GE scanner in the Advanced Magnetic Imaging Centre (AMI-Centre) of the Helsinki University of Technology.

It is common that under a hypothesis driven experiment the scanning is focused only on a few slices of the brain, which have been classified as interesting beforehand. Sometimes this is beneficial, since it could allow faster scanning or increased resolution, but as ICA is a purely data-driven method it would be rather impossible to define the interesting regions of the brain beforehand. And again, such scanning would limit the data too much to fully characterize unexpected phenomena.

6.1.3 Volume Preparation

Before analysis, the volumes were processed in the usual way (*c.f.*, Worsley and Friston, 1995) fMRI data is processed when using the traditional analysis method (see Section 2.3.2). This was done with the SPM toolbox (SPM, 1999) and resulted, for each of the 14 subjects, in 80 volumes with a resolution of $95 \times 79 \times 69$ voxels. Additionally, the volumes were masked with a cortical mask to remove uninteresting voxels outside the brain. This effectively lowered the amount of data to half, still leaving an intimidating 80×254484 observation matrix per subject.

6.2 Individual Experiments

The observation matrix of each subject was then analyzed with the method described in Section 4.3. The FastICA algorithm was run 100 times using a bootstrapping of 20% and PCA whitening to the 30 strongest principal components in each run. FastICA was used in *symmetric* mode with nonlinearity *tanh*, estimating 15 independent components in each run.

This resulted in 1500 independent component estimates per subject, which were then clustered with correlation threshold 0.8 and power 8. All parameter values in the experiments were chosen heuristically, based on earlier trials.

6.3 Group Experiment

As mentioned earlier the described individual experiment was repeated under the same conditions for all 14 subjects. Although the analysis of consistency within the whole group is beyond the scope of this thesis, some general observations can still be possible and could prove valuable. Therefore the complete results for all subjects were kept in an easily comparable form.

Chapter 7

Results

7.1 Individual Results

Because of the large amount of subjects and since showing the volumetric activation patterns of the independent components on paper can be problematic, this chapter highlights only the most interesting and surprising results. The complete results for all individual experiments are shown in Appendix B. Although similar components are found on many subjects, as discussed in Section 7.2, the results are shown using the clearest examples. The visualization is explained in detail in Chapter 5 and the information on the human brain, needed to understand the medical explanations, is in Chapter 2.

7.1.1 Overview

Overall, the individual results shown in Figures B.1 – B.14 are very good, and clearly demonstrate the usability of the analysis method and visualization tools. They are easy and reliable to interpret. The results show that the components closely related to the stimulus, explained in Chapter 6, are very consistent. There are also many other consistent components that have either an interesting spatial or temporal structure, or both. Additionally, some of the less consistent components are still interesting, perhaps revealing surprising phenomena.

Particularly, the results for subjects JK, PK and SN seem to contain many interesting components related to the stimulus. The majority of those components are also highly consistent. On the other hand, there seems to be a

very strong scanning anomaly in the data of subject TL. Also, a similar phenomenon can be seen in the results for subject MG, but in that case the artifact appears to contaminate the whole results.

Note how the measures of discrimination, variability and skewness of a component together seem to reveal how reliable, and possibly interesting, the component is. The results are interesting both from a theoretical point of view, validating the method, and from a medical point of view, possibly revealing unexpected relations within human brain processing of speech and language.

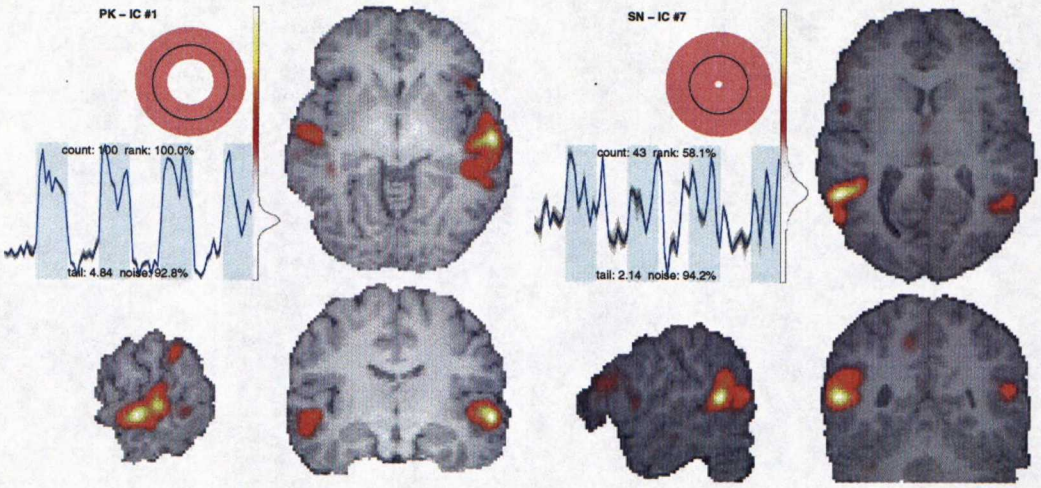
7.1.2 Components Related to Stimulus

Components that are closely related to the stimulus are quite predictable and should be very consistent with all analysis methods. For example, Figure 7.1 shows two interesting components, the first from subject PK and the second from subject SN. Similar components can be found practically from all subjects, but these are shown just as two very clear cases. The first component appears to correspond to activation on the primary auditory areas, and mainly on the right hemisphere. The second component seems to relate to a secondary auditory area, sometimes referred to as Wernicke's area. The activation on the second component is clearly stronger on the left-hand side. The figure shows only the three slices centered on the point of strongest activation, but even without the possibility of interactively studying the whole volumes, the skewed histograms support the fact that the volumes contain very strong and focal activation.

Both components are also very consistent with time-courses strongly related to the stimulus, with the first component being almost a perfect match. This is predictable, since the activation on auditory areas should follow the speech stimulus quite closely. Additionally, the discrimination of the components is very good, that is, they are well separated from other components. The disks depicting the intra-group distances are mere dots and would easily fit inside the inter-group disks, although the inter-group distances have a somewhat bigger spread on the second case.

7.1.3 Components Revealing Artifacts

Two examples of clear artifacts are shown in Figure 7.2. The first is from subject JK and related to filtering, and the second reveals a very strong scanning anomaly in the first scan from subject TL. Both components are well discrimi-



(a) (b)

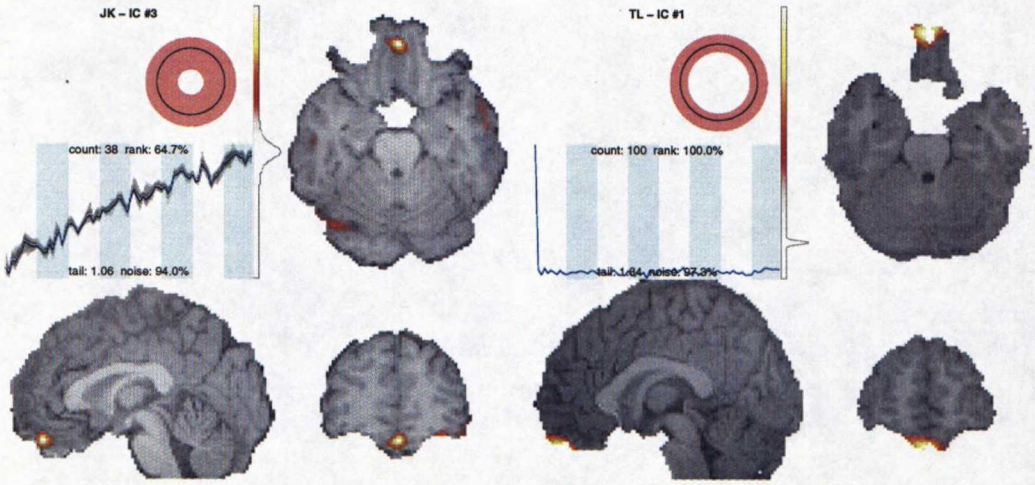
Figure 7.1: Closer look at consistent stimulus related components. The (a) primary auditory areas and (b) secondary auditory areas.

nated and very consistent. Considering the complete results for these subjects, see Figures B.4 and B.11 respectively, it is clear that ICA is able to remove the artifacts from the rest of the data by isolating them as separate components. Although the multiple runs of ICA, explained in Chapter 4, are able to separate the filtering related artifact, the spread of inter-group distances is still quite large. Clearly, the artifact that affects only a single time point is easier to separate.

The histogram of the first example is very symmetric, suggesting that there should not be any strong focal activation in that volume. Actually, a closer look at the whole volume reveals it to be completely covered by a pattern that resembles ripples, like often seen on the surface of water. The fast fMRI scanners often produce such artifacts, and the signals are actually filtered to remove such patters. However, it seems that the phenomenon causing the artifact is not fully stationary, and the simple constant filtering still leaves part of the contamination with an almost linearly drifting time-course. This new information, revealed by ICA, could be used to improve the filters used in the scanning.

On the other hand, the histogram of the second artifact is very skewed and the time-course shows that the artifact is only present in the first volume of the fMRI sequence. This artifact seems to result from a strong ghosting or shadowing phenomenon located just below the frontal lobe of the brain. It is almost completely outside the volume used in the analysis, and only partially

visible in the example slices.



(a) (b)

Figure 7.2: Closer look at clear aftifact components. (a) A filtering artifact and (b) a scanning anomaly in the first volume of the sequence.

7.1.4 Components with Strong Variability

The examples in Figure 7.3 show the true benefits of the multiple run method. These cases are more complex than the previous ones, and would practically be impossible to analyze reliably using only a single run of ICA, or the traditional SPM method explained in Section 2.3.3. Both examples show a component with strong variability, yet, the nature of the variability is very different in each case.

The component in Figure 7.3(a), from subject KR, seems to contain activation on the visual area. The activation pattern itself seems rather clear with a relatively skewed histogram, much like the auditory examples in Figure 7.1. Additionally, activation on the visual cortex is common in all kinds of studies, since the eyes provide a constant flow of information, even when the subject would concentrate on listening. However, the strong temporal variability suggest that the component is not very consistent, or reliable. Indeed, focusing on the disks showing the discrimination measures reveals that the intra-group distances are very small, excluding an outlier, but the inter-group distances are more uniformly spread and the discrimination is not very good. This means that the component estimation itself may be stable, but the component is very difficult to separate from other components. One explanation for such

behavior could be that the activation on the visual area may be too weak, or unstructured, to detect reliably under the auditory stimulation.

On the other hand, the situation in Figure 7.3(b), which shows a component from subject SN, is very different. The highly symmetric histogram suggests that the activation pattern does not contain very strong points. Still, there seems to be some focal patterns around the brain stem. In this case, the small spread and large mean of the inter-group distances suggest that the component is quite well separated from other components. The estimation of the component itself is much more unstable than that in the previous case, since the mean value of the intra-group distances is quite big, and there is actually a hole in the middle of the disk. Also, the very high number of grouped estimates supports the interpretation that the component is able to identify a distinct, but very unstable, phenomenon. Most probably, the component reveals a portion of the main blood vessels, rising around the brain stem. Since the fMRI scanning is too slow to accurately follow the flow of oxygenated blood in the vessels, ICA could detect a separate signal subspace, but seems to be unable to accurately decompose that subspace into independent components. Additionally, the time-course of the component is very structured, although highly varied, and could be related to the heart beat.

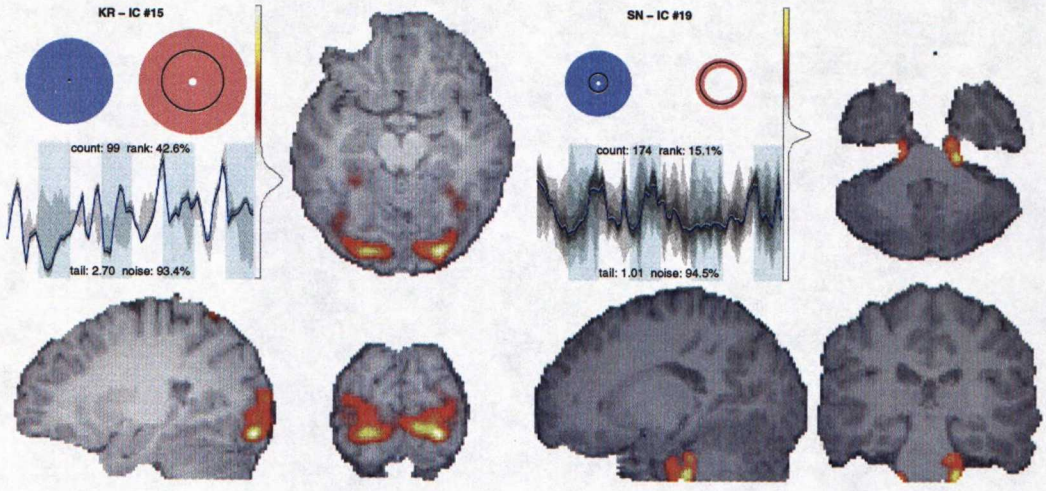


Figure 7.3: Closer look at components with strong variability. (a) Activation on the visual area and (b) a possible portion of the blood vessels.

7.1.5 Other Interesting Components

Two other interesting examples are shown in Figure 7.4. The first from subject PK, showing potential activation in the Broca's area, and the other, from subject JK, which reveals activation around the frontal cingulate gyrus. The components are very consistent, and the time-courses suggest that they are not completely unrelated to the stimulus. These components are interesting because Broca's area is involved in the production of sentences and areas around the cingulate gyrus may relate to long-term memory. Activation around these areas while listening to spoken language is quite reasonable. Still, the components are so weakly related to the stimulation that they would be very difficult to detect using the traditional SPM method.

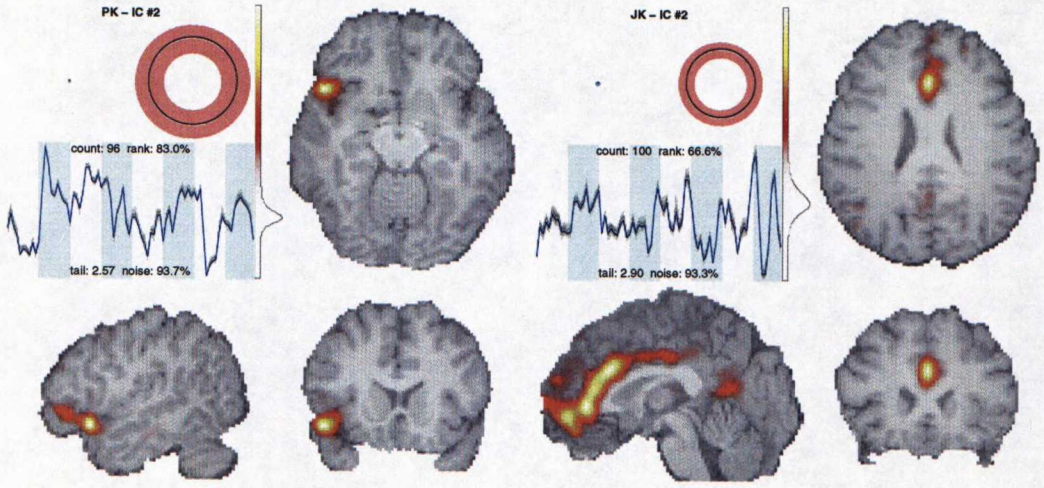


Figure 7.4: Closer look at some of the other interesting components. Activation possibly on (a) the Broca's area and (b) frontal cingulate gyrus.

7.1.6 Spatial Relations Revealed by Variability

As mentioned in Chapter 4, the variability of a component can also be analyzed spatially by calculating the spread of the group using the volumes. Although the computation using the whole volumes is quite demanding, the resulting volumes of spatial variability can be used just like the temporal spread around the mean time-courses.

A very interesting example is shown in Figure 7.5, from subject SN, where the spatial variability reveals a relation that can not be seen from the time-courses

alone. A slice of component 10 is shown with that of component 7. The second slice of component 10, shown in the middle, is the spatial variability. The spatial pattern of component 7 is remarkably similar to the variability of component 10. This suggests that the main source of variability for component 10 is actually component 7, that is, the estimation of component 10 tends to, from time to time, get mixed with component 7. The components may, for example, span a subspace together.

(a) IC #10 (b) VAR #10 (c) IC #7

Figure 7.5: Spatial variability clearly reveals related components from subject SN. Horizontal slices of (a) component 10 and (b) its spatial spread next to (c) component 7. The main source of variability for component 10 is clearly component 7.

7.2 Group Results

Although the scope of the thesis does not include group analysis, the amount of information provided by the method allows quite fast and reliable human comparison of the subjects. The similarities, and differences, among the subjects are easy to spot with the help of the variabilities. Some of the components, shown only as horizontal slices in Appendix B, may be quite hard to identify without the possibility to interactively study the whole volumes. On the other hand, the task is not simple as similarities can be identified in many ways. Subjects may share a spatially matching activation pattern with possible differences in the time-courses or the time-courses may be similar with some differences in the spatial patterns. In some cases, both may be very similar.

Some of the components shared by multiple subjects are summarized in Table 7.1. The components used in the table are (PA) the primary auditory areas, as in Figure 7.1(a); (BV) the blood vessels around the brain stem in Figure 7.3(b); (AA) secondary auditory areas, as in Figure 7.1(b); (C) the areas around the cingulate gyrus, as in Figure 7.4(b); (VE) large ventricles

located in the middle of the brain; (VI) the visual area in Figure 7.3(a); and (F) the filtering artifact in Figure 7.2(a).

The comparison suggests that the subjects can be categorized into three groups. In the first, the majority of the subjects share many of the components Second, some subjects seem to have only a subset of the components, which may be caused by, for example, individual differences or level of attention. And third, a few subjects share almost no similarities with the others. This may be explained by the fact that the data of these few subjects seems to be very heavily contaminated by scanning artifacts.

Subject	Figure	Components						
		PA	BV	AA	C	VE	VI	F
SN	B.10	●	●	●	●	●	●	●
KR	B.5	●	●	●	●	●	●	●
TL	B.11	●	●	●	●	●	○	●
MT	B.7	●	●	●	●	●	●	○
JK	B.4	●	●	●	●	○	○	●
RS	B.9	●	●	●	●	○	○	●
TT	B.13	●	●	●	●	○	●	○
HH	B.2	●	●	●	○	●	●	○
HR	B.3	●	●	●	○	●	●	○
PK	B.8	●	●	●	○	●	○	○
AS	B.1	●	○	○	●	○	●	●
UL	B.14	●	●	○	○	○	○	●
TP	B.12	●	●	○	●	○	○	○
MG	B.6	●	○	○	○	●	○	○

Table 7.1: Consistency across subjects shown as a table of very similar components shared by multiple subjects. Definitions of the components: PA is the primary auditory areas, BV is the blood vessels around the brain stem, AA is secondary auditory areas, C is the cingulate gyrus, VE is the main ventricles, VI is the visual area and F is the filtering artifact.

Chapter 8

Conclusions

8.1 Discussion

A method for analyzing the consistency of independent components was presented and its usefulness was tested in a real fMRI study. Although the method was used with FastICA, it should be relatively straightforward to use it with other ICA and BSS algorithms.

The method works by exploiting the variability of ICA algorithms in a bootstrapping and clustering based approach. The method produces important additional information on the estimated independent components, which helps in interpreting the results, and it allows the detection of sources difficult or impossible to find with ICA alone or even with other bootstrapping methods.

The experiments show that the method works well with real fMRI data and, indeed, makes interpreting the results easier and more reliable. The method verifies the consistency of the expected results, but also reveals unexpected components in a reliable way. Additional information on the variability of the less consistent components helps interpreting the underlying phenomena.

Additionally, the visualization tools are able to give fast and clear overviews of the analysis results, but also allow the indepth study of the most interesting features. This makes interpreting the results even easier.

8.2 Medical Relevance

Consistent analysis of fMRI studies is crucial in the medical field and the presented method offers a way to do exactly that. The additional information on the variability, especially the location, is also interesting. Possibly helping to improve other methods and form new hypotheses on brain function as well.

The added reliability of the interpretations made from results produced with the method is very important from a medical point of view. The proposed method is also considerably faster that other similar methods and makes it more usable in real medical research.

8.3 Future Research

The presented method and tools will be used extensively on future research projects. This could allow further improvements to be made on the clustering and even on running ICA multiple times, when additional experience is gathered. The visualization tools could be made more compatible with other medical tools allowing better integration into existing analysis setups.

The additional information on the source and nature of variability could also be used to improve the original solution (*c.f.*, Friman et al., 2004). The framework of denoising source separation (DSS) (Särelä and Valpola, 2005) allows the use of such information in a highly integrated way during the estimation. Additionally, in DSS the estimation does not have to be based strictly on statistical independence.

When variability with different natures can be identified and analyzed, there is no reason why the method presented could not be extended also for group studies, where a similar problem exists in comparing results across many subjects (*c.f.*, Calhoun et al., 2001a,b, McNamee and Lazar, 2004). Solutions to the problems in group studies are important, since group studies are standard in the medical field.

Appendix A

Mathematical Concepts

A.1 Principal Component Analysis

Principal component analysis (PCA) is a powerful statistical method for multivariate data analysis. It is very efficient and simple to use for noise reduction, feature selection, compression, decorrelation and whitening. It is frequently used in practically all signal processing. Consider reading Haykin (1998), Jolliffe (2002) for more information on principal component analysis and algorithmic implementations.

A.1.1 Eigenvalue Decomposition

The eigenvalues λ and corresponding eigenvectors $\mathbf{v}$ of an $n \times n$ square matrix $\mathbf{M}$ are defined as the solutions of:

$$\mathbf{M}\mathbf{v}_i = \lambda_i \mathbf{v}_i \ , \tag{A.1}$$

where $i = 1, \ldots, n$. The rank of matrix $\mathbf{M}$ defines how many non-zero eigenvalues there are. Another way of expressing the same solutions is the eigenvalue decomposition (EVD):

$$\mathbf{M} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1} \ , \tag{A.2}$$

where the columns of matrix $\mathbf{V}$ contain the eigenvectors and $\mathbf{D}$ is a diagonal

matrix of the corresponding eigenvalues. Replacing a matrix with its EVD is very practical, since many mathematical operations are simpler and faster to perform on the decomposition. Additionally, the structure of the matrix is easy to study from its EVD.

A.1.2 Principal Components

The principal components of multivariate data $\tilde{\mathbf{X}}$ are defined as the orthonormal basis vectors, which contain the maximum variance of the data. Each row of the data matrix $\tilde{\mathbf{X}}$ contains a different observation with samples on the columns. One way of finding the principal components is to use the EVD of the estimated covariance matrix of the data. First of all, the covariance matrix $E\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T\}$, where $\tilde{\mathbf{x}}$ is a column of $\tilde{\mathbf{X}}$ and the expectation function $E\{\cdot\}$ simply normalizes the values, is a real symmetric matrix. Therefore, the following must hold for the EVD of a covariance matrix:

$$\mathbf{M}^T = \mathbf{V}^{-1^T}\mathbf{D}^T\mathbf{V}^T = \mathbf{V}\mathbf{D}\mathbf{V}^{-1} = \mathbf{M} \ . \tag{A.3}$$

Clearly, this is only possible when $\mathbf{V}^{-1} = \mathbf{V}^T$, thus the eigenvectors must be orthogonal. Additionally, assuming the data has zero mean and the eigenvectors are normalized, that is, $E\{\tilde{\mathbf{x}}\} = 0$ and $||\mathbf{v}|| = 1$, the following holds for the data projected along a principal direction ${\mathbf{v}_i}^T\tilde{\mathbf{x}}$:

$$\begin{aligned} &E\{{\mathbf{v}_i}^T\tilde{\mathbf{x}}\} = {\mathbf{v}_i}^T E\{\tilde{\mathbf{x}}\} = 0 \ , \text{ and} \\ &{\sigma_i}^2 = E\{({\mathbf{v}_i}^T\tilde{\mathbf{x}})^2\} = {\mathbf{v}_i}^T E\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T\}\mathbf{v}_i = \lambda_i \ . \end{aligned} \tag{A.4}$$

Thus, the eigenvalues corresponding to the orthonormal eigenvectors equal the variances of the data projected along the eigenvector directions. Therefore, decomposing the covariance matrix with EVD gives the principal components of the data. The usual convention is to make the one with the largest eigenvalue, or variance, the first principal component and so on.

A.1.3 Whitening

Whitening observed data $\tilde{\mathbf{X}}$, that is, making the rows uncorrelated and their variances equal to unity, often makes further processing easier. Whitening can be expressed as a linear transformation and one way to find such a transfor-

mation is based on the EVD of the estimated covariance matrix of the data. Again, assuming the data has zero mean, this equals:

$$E\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T\} = \mathbf{V}\mathbf{D}\mathbf{V}^T \ . \tag{A.5}$$

The following whitening transformation can be constructed from the eigenvalues and eigenvectors of the decomposition:

$$\mathbf{X} = \mathbf{D}^{-1/2}\mathbf{V}^T\tilde{\mathbf{X}} \ , \tag{A.6}$$

where the operations on the diagonal matrix $\mathbf{D}$ can be calculated simply component-wise. Using the columns $\mathbf{x}$ of the whitened data matrix $\mathbf{X}$ it is easy to check, straight from Equations A.5 and A.6, that $E\{\mathbf{x}\mathbf{x}^T\} = \mathbf{I}$. For the following ICA, this means that the whitening effectively transforms the mixing matrix into a new one. When the original ICA model is defined as $\tilde{\mathbf{X}} = \tilde{\mathbf{A}}\mathbf{S}$, it can be seen from Equations 3.1 and A.6 that:

$$\mathbf{X} = \mathbf{A}\mathbf{S} = \mathbf{D}^{-1/2}\mathbf{V}^T\tilde{\mathbf{A}}\mathbf{S} \ . \tag{A.7}$$

What makes this transformed model useful is that the new mixing matrix $\mathbf{A}$ is orthonormal, which reduces the number of free parameters. This is evident from:

$$E\{\mathbf{x}\mathbf{x}^T\} = \mathbf{A}E\{\mathbf{s}\mathbf{s}^T\}\mathbf{A}^T = \mathbf{A}\mathbf{A}^T = \mathbf{I} \ , \tag{A.8}$$

where the independent components $\mathbf{S}$ are also assumed white. This is not a very restricting assumption, since they can be later transformed with the inverse of the whitening transformation anyway. Actually, it is shown, later on, that independence implies uncorrelation. Therefore, assuming that $E\{\mathbf{s}\mathbf{s}^T\} = \mathbf{I}$ is merely a scaling issue, which causes no problems, since the scaling in ICA is ambiguous anyway.

A.1.4 Reducing Dimension

Additionally, using the EVD of the covariance matrix to form the whitening transformation allows the reduction of free parameters also in another way.

By selecting only a subset $k \leq n$ of the eigenvalues, and corresponding eigenvectors for the transformation, the dimension of the whitened data can be lowered effectively. The dimension reduction also reduces noise, that is, improves the signal-to-noise ratio. Additive noise, for example Gaussian, often appears distributed along several directions with a small energy. If the selected eigenvalues are the largest ones, which identify the directions with the highest variances, the dimension reduction is optimal in the sense that it retains as much of the original signal power as possible.

A.2 Estimating Independence

The orthogonal principal components identified with PCA and, therefore, also the whitened data vectors are uncorrelated, which means that their covariance is zero:

$$E\{v_i v_j\} - E\{v_i\}E\{v_j\} = 0 \ . \tag{A.9}$$

Uncorrelated vectors are sometimes, confusingly, called linearly independent. However, true statistical independence means that the joint probability density function (pdf) $p(\mathbf{s})$ is factorisable on its marginal probability densities as:

$$p(\mathbf{s}) = p(\mathbf{s}_1, \ldots, \mathbf{s}_k) = \prod_{i=1}^{k} p(\mathbf{s}_i) \ . \tag{A.10}$$

This means that the marginal distributions contain no information on each other whatsoever. It also means that statistical independence is a stronger restriction than uncorrelation, since the following holds for arbitrary functions $f_i(\cdot)$ and $f_j(\cdot)$:

$$E\{f_i(\mathbf{s}_i)f_j(\mathbf{s}_j)\} = E\{f_i(\mathbf{s}_i)\}E\{f_j(\mathbf{s}_j)\} \ . \tag{A.11}$$

Additionally, as directly seen from Equations A.9 and A.11, independence implies uncorrelation. This is actually the reason why whitening the data before ICA will not alter the independent components. The following introduces the key concepts in approximating statistical independence, but for more information consider reading Hyvärinen and Oja (2000), Cardoso (2003).

Since a closed form solution would require exact determination of the pdfs, which is generally not possible, the independence has to be approximated. A good starting point is the entropy H of a discrete random variable $\mathbf{s}$, defined as:

$$H(\mathbf{s}) = -\sum_i p(\mathbf{s} = a_i) \log p(\mathbf{s} = a_i) \ , \tag{A.12}$$

where the a_i are all the possible values of $\mathbf{s}$. By definition, entropy measures the amount of information contained in the random variable $\mathbf{s}$. This corresponds to the definition of differential entropy for continuous random variables.

A.2.1 Mutual Information

Mutual information measures how much information is shared among random variables. The mutual information I between k random variables is defined using entropy as:

$$I(\mathbf{s}_1, \ldots, \mathbf{s}_k) = \sum_{i=1}^{k} H(\mathbf{s}_i) - H(\mathbf{s}) \ . \tag{A.13}$$

Clearly, mutual information between independent variables should be 0. Thus, estimating the independent components is possible by minimizing the mutual information between them. However, doing this in practice can be computationally very heavy.

A.2.2 Negentropy

Gaussian variables have the largest entropy among all random variables of equal variance. Negentropy J is defined as:

$$J(\mathbf{s}) = H(\mathbf{s}_{gauss}) - H(\mathbf{s}) \ , \tag{A.14}$$

where $\mathbf{s}_{gauss}$ is a Gaussian random variable with the same covariance as $\mathbf{s}$. The important point is that as $\mathbf{s}$ is considered to be white, negentropy differs from

mutual information only by a constant:

$$I(\mathbf{s}_1, \ldots, \mathbf{s}_k) = C - \sum_{i=1}^{k} J(\mathbf{s}_i) \ . \tag{A.15}$$

Therefore, maximizing negentropy equals minimizing mutual information when estimating independence. Although negentropy is computationally simpler than mutual information, it is also based on the exact pdfs. To make the estimation feasible in practice, the requirement of knowing the exact pdfs should be eliminated. Fortunately, it turns out that negentropy can be approximated without knowing the exact pdfs.

A.2.3 Non-Gaussianity

Since negentropy is essentially measuring the difference between the Gaussian distribution and that of the independent variables, it can be approximated by estimating the non-Gaussianity of the random variables directly. Classical measures of non-Gaussianity are skewness and kurtosis, or the third and fourth order cumulants. The kurtosis of s is defined as:

$$kurt(s) = E\{s^4\} - 3E\{s^2\}^2 \ . \tag{A.16}$$

Actually, as s is assumed to have unit variance, this simplifies to $E\{s^4\} - 3$. Kurtosis is zero for a Gaussian random variable, negative for sub-Gaussian and positive for super-Gaussian. Thus, the absolute value of kurtosis can be used as a measure of non-Gaussianity. Maximizing the norm of kurtosis, which is roughly equivalent to maximizing independence, is computationally very efficient, making ICA feasible in practice.

A very precise estimation of non-Gaussianity can be constructed as a weighted sum of both skewness and kurtosis, but it is often decided to use only one of them in practice. The possibility of using non-Gaussianity as a measure of independence is not so surprising when considering the central limit theorem. Basically, it states that the distribution of a mixture of *i.i.d.* random variables tends to be more Gaussian than the original ones. Therefore, the more non-Gaussian a variable is the more independent it has to be. Incidentally, this link between independence and non-Gaussianity is the reason why only one of the components in ICA can originally be Gaussian.

A.3 Fast Fixed-Point Iteration

One of the most widely used implementations of ICA is called FastICA. It uses a Newton-iteration based fixed-point optimization scheme and an objective function based on negentropy. FastICA can search for the independent components one at a time, in *deflation* mode, or *symmetrically*, all at once. Its performance can also be tuned somewhat by choosing the nonlinearity used in the estimation, for example between the *cubic* or the *tanh*. For more information on the FastICA algorithm, and its derivation, consider reading Hyvärinen and Oja (1997, 2000). For the actual implementation see FastICA (1998).

A.3.1 Deflation Mode

The independent components are estimated one by one and the iteration simply prevents the estimation of the ith component from converging to the same solution as any of the previous $1, \ldots, i-1$ components. The basic FastICA iteration takes the following form in deflation mode:

1. Begin with an initial (*e.g.* random) weight vector $\mathbf{w}_i$.
2. Let $\mathbf{w}_i{}^+ = E\{\mathbf{x}g(\mathbf{w}_i{}^T\mathbf{x})\} - E\{g'(\mathbf{w}_i{}^T\mathbf{x})\}\mathbf{w}_i$
3. Let $\mathbf{w}_i{}^+ = \mathbf{w}_i{}^+ - \sum_{j=1}^{i-1} \mathbf{w}_i{}^{+T}\mathbf{w}_j\mathbf{w}_j$
4. Let $\mathbf{w}_i = \mathbf{w}_i^+ / ||\mathbf{w}_i^+||$
5. Iterate steps 2. – 4. until convergence.

Since the signs of the independent components are not fixed, convergence means that the old and new values of $\mathbf{w}$ are parallel. The function $g(\cdot)$, and its derivative $g'(\cdot)$, in step 2. is the selectable nonlinearity and using the *cubic* actually corresponds to estimating the kurtosis.

A.3.2 Symmetric Mode

In deflation mode, the previously estimated components are priviledged in step 3., that is, only the currently estimated weight vector is affected, and the estimation order matters. For example, estimation error accumulates into the last components. In symmetric mode, the order in which the components are

estimated has no effect. This is archieved by estimating all the $\mathbf{w}$, that is, the whole matrix $\mathbf{W}$, first and orthogonalizing the matrix only after that as an additional step.

Appendix B

Complete Results

B.1 Description

The following figures show the complete results of the individual experiments. The details of the experiments are explained in Chapter 6 and the most important findings are discussed in Chapter 7. More information on the visualization used in the figures can be found in Chapter 5 and also in Chapter 7. A quick reference of the figures for different subjects, and pages they are on, is shown in Table B.1.

B.2 Figures

Subject	Figure	Page	Subject	Figure	Page
AS	B.1	64	PK	B.8	71
HH	B.2	65	RS	B.9	72
HR	B.3	66	SN	B.10	73
JK	B.4	67	TL	B.11	74
KR	B.5	68	TP	B.12	75
MG	B.6	69	TT	B.13	76
MT	B.7	70	UL	B.14	77

Table B.1: Quick reference of the figures for different subjects.

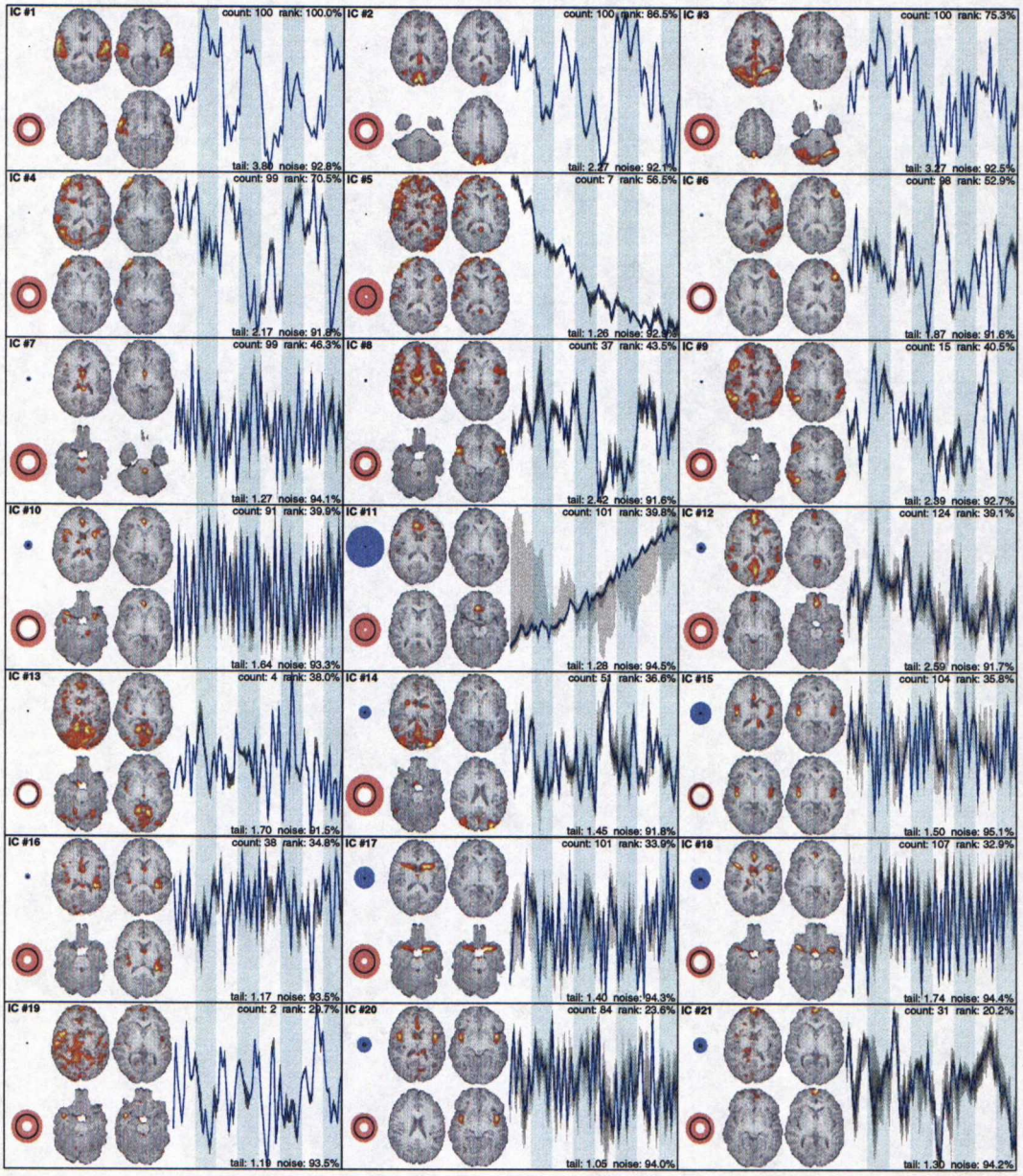


Figure B.1: Results for subject AS. Each independent component is visualized with grouping and consistency information. The disks on the left show the spread of the cluster size and distance to other clusters. The four horizontal slices of the brain show the activation pattern from different heights. Also, the activation time-course is shown with the quantiles as different shades of gray.

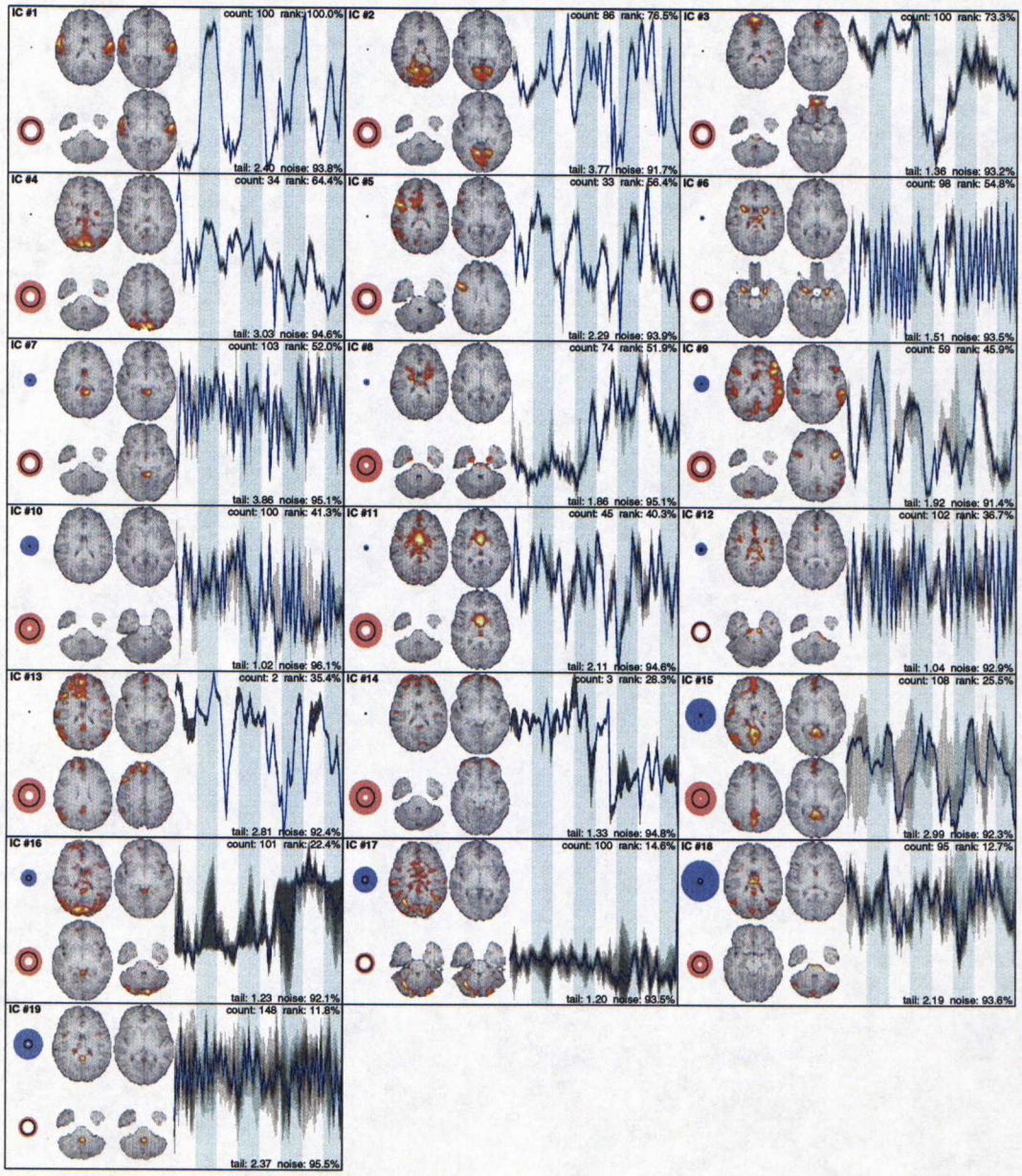


Figure B.2: Results for subject HH. Each independent component is visualized with grouping and consistency information. The disks on the left show the spread of the cluster size and distance to other clusters. The four horizontal slices of the brain show the activation pattern from different heights. Also, the activation time-course is shown with the quantiles as different shades of gray.

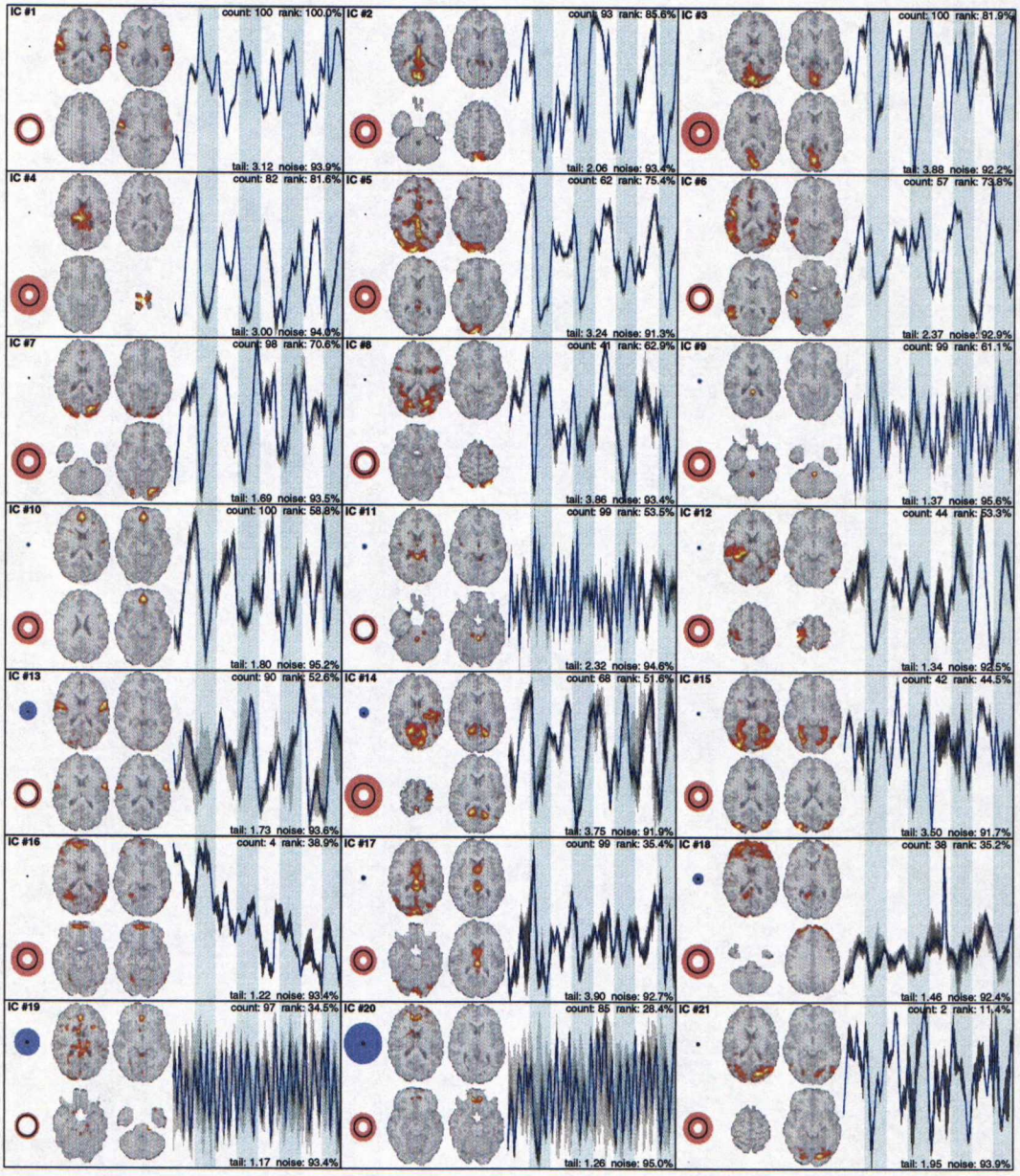


Figure B.3: Results for subject HR. Each independent component is visualized with grouping and consistency information. The disks on the left show the spread of the cluster size and distance to other clusters. The four horizontal slices of the brain show the activation pattern from different heights. Also, the activation time-course is shown with the quantiles as different shades of gray.

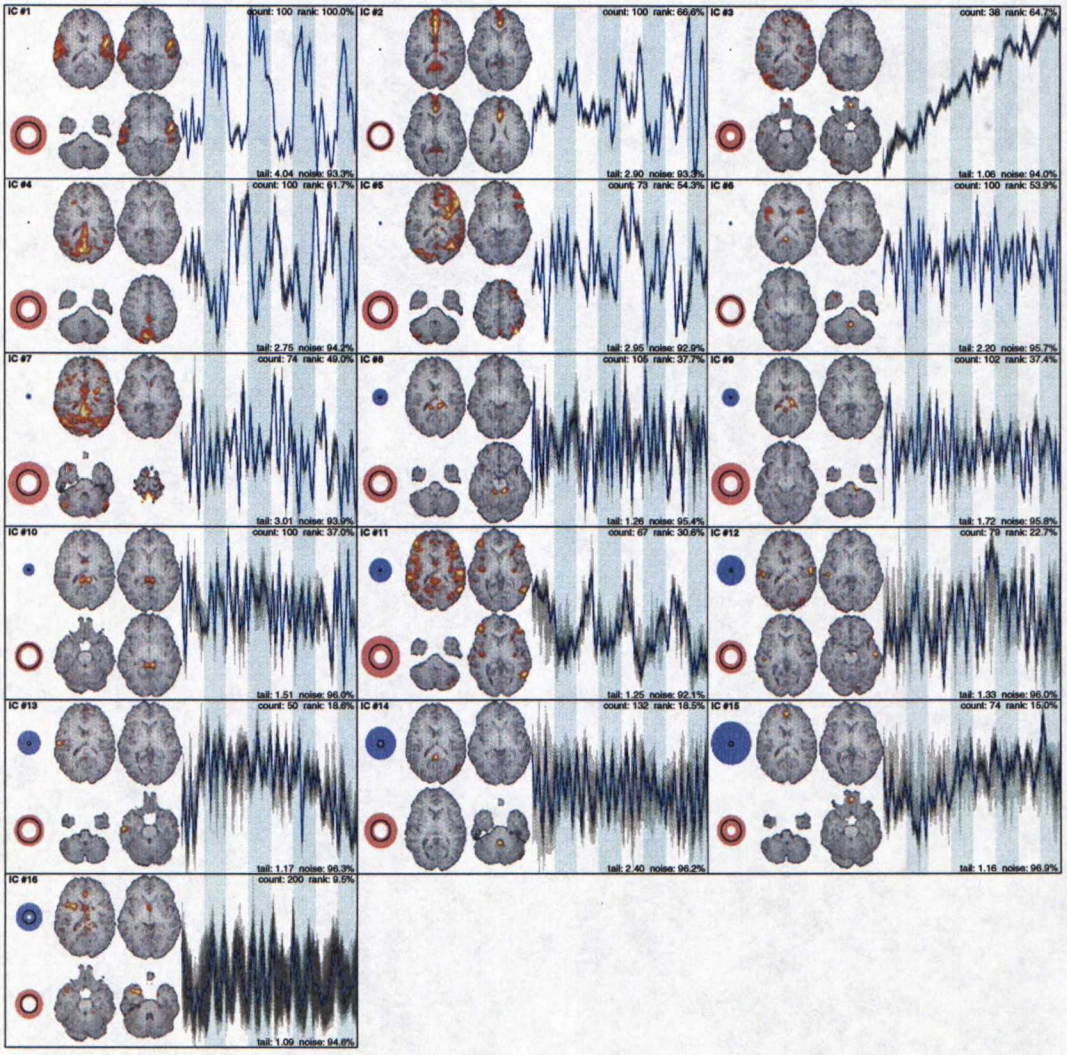


Figure B.4: Results for subject JK. Each independent component is visualized with grouping and consistency information. The disks on the left show the spread of the cluster size and distance to other clusters. The four horizontal slices of the brain show the activation pattern from different heights. Also, the activation time-course is shown with the quantiles as different shades of gray.

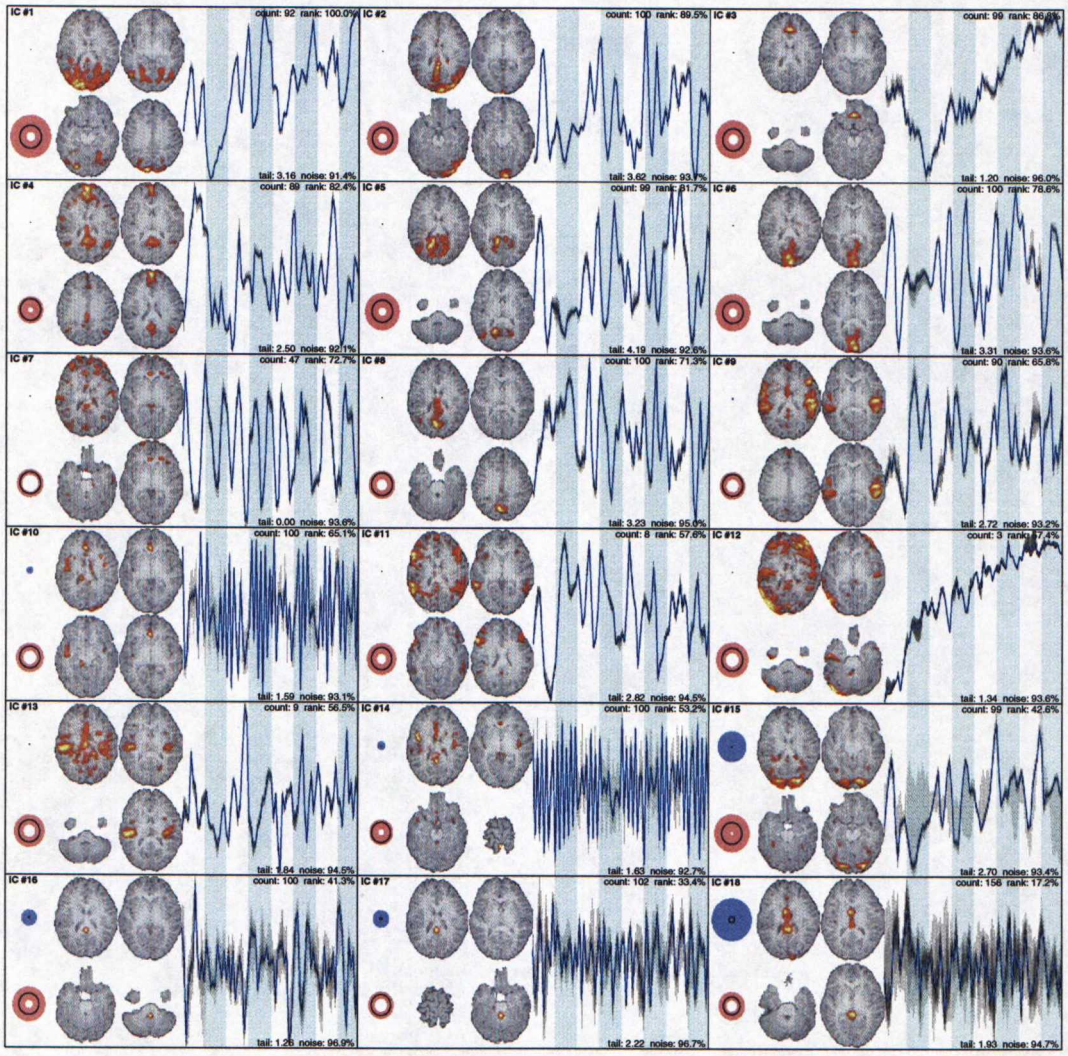


Figure B.5: Results for subject KR. Each independent component is visualized with grouping and consistency information. The disks on the left show the spread of the cluster size and distance to other clusters. The four horizontal slices of the brain show the activation pattern from different heights. Also, the activation time-course is shown with the quantiles as different shades of gray.

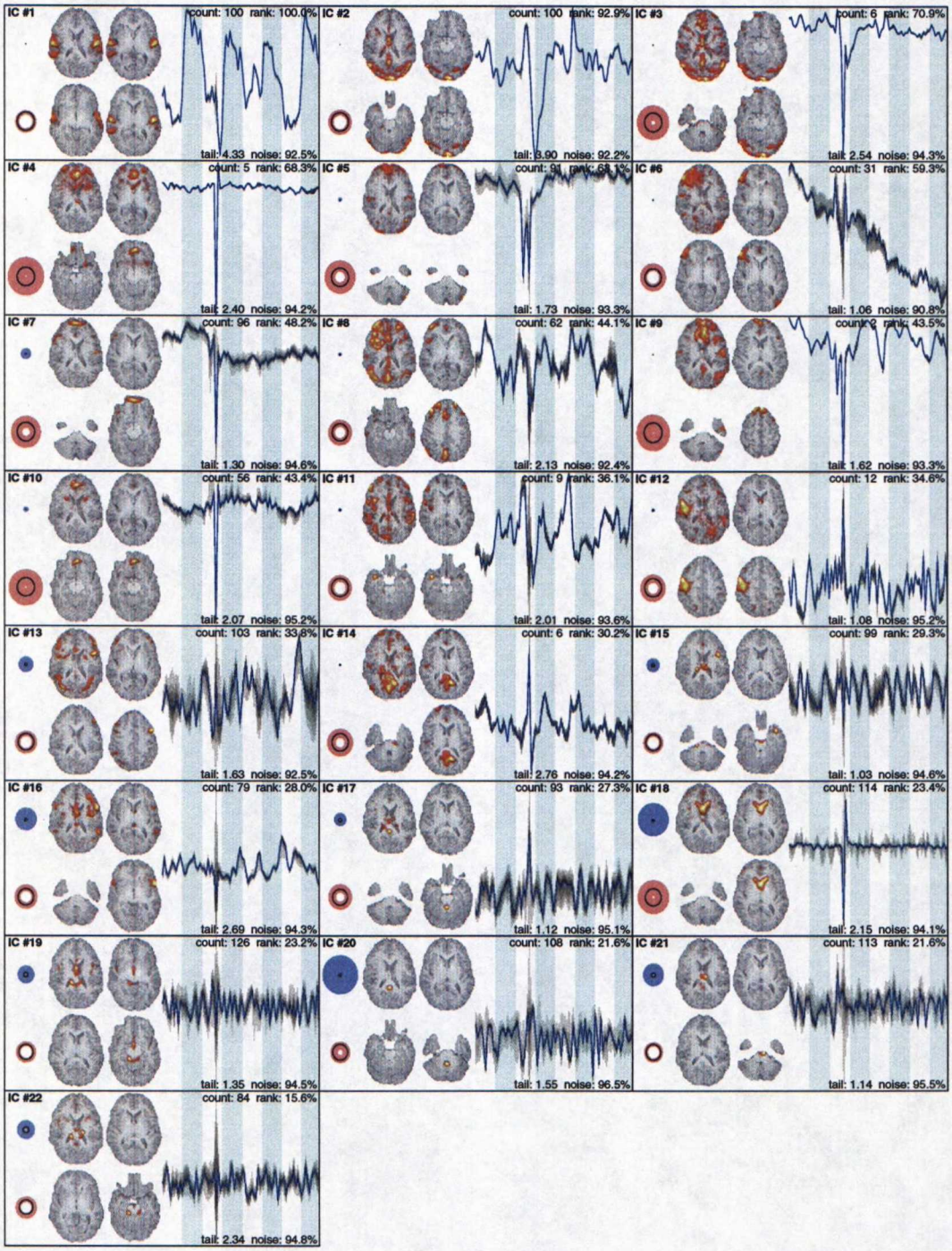


Figure B.6: Results for subject MG. Each independent component is visualized with grouping and consistency information. The disks on the left show the spread of the cluster size and distance to other clusters. The four horizontal slices of the brain show the activation pattern from different heights. Also, the activation time-course is shown with the quantiles as different shades of gray.

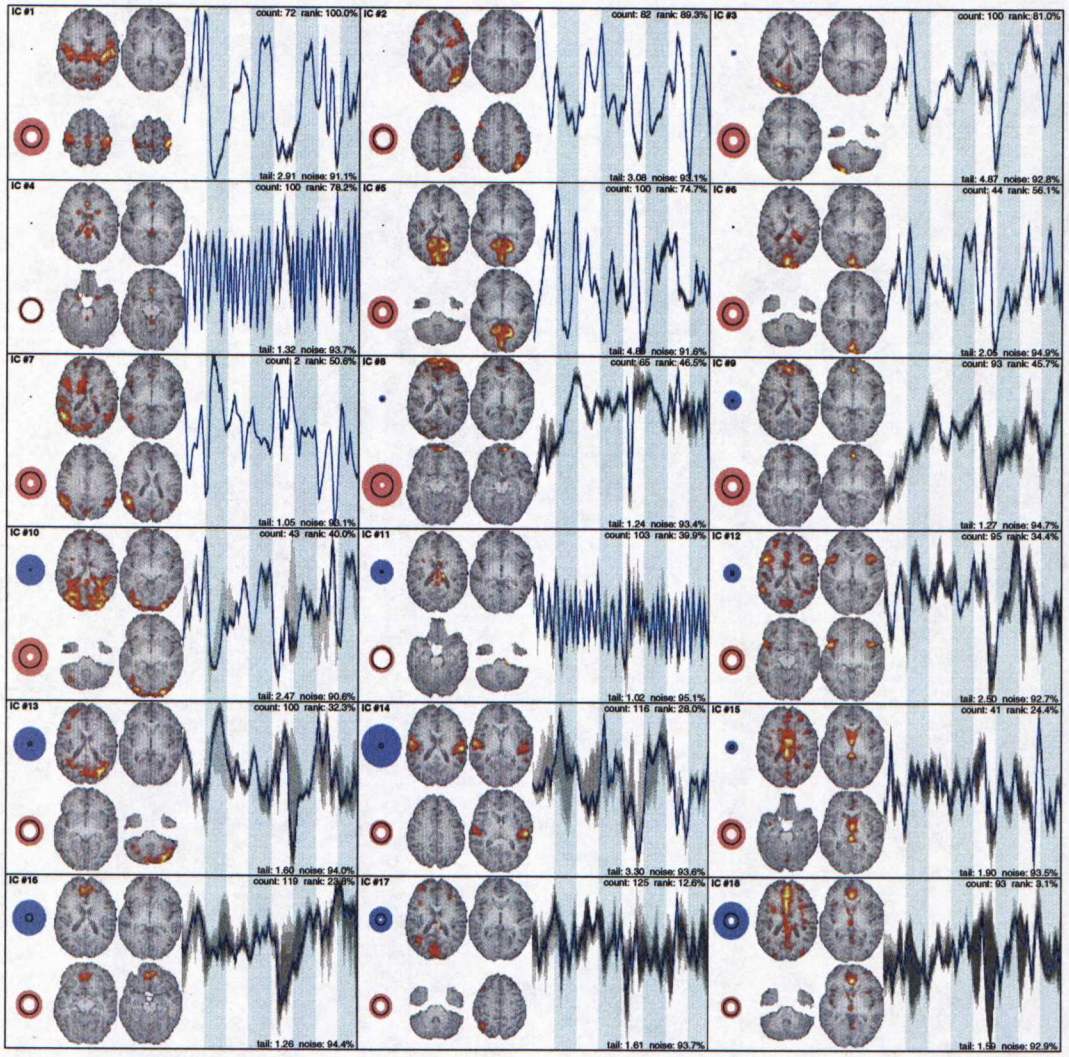


Figure B.7: Results for subject MT. Each independent component is visualized with grouping and consistency information. The disks on the left show the spread of the cluster size and distance to other clusters. The four horizontal slices of the brain show the activation pattern from different heights. Also, the activation time-course is shown with the quantiles as different shades of gray.

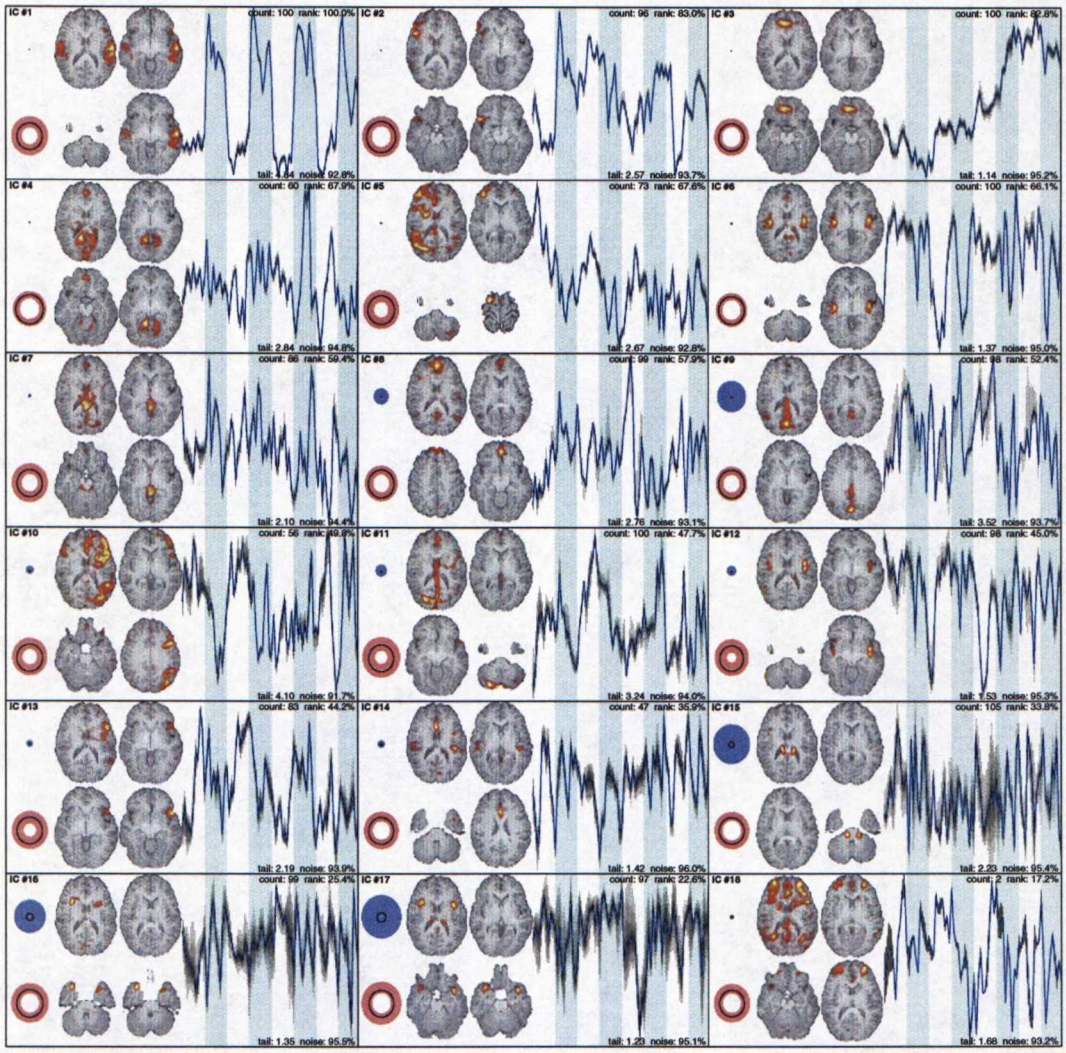


Figure B.8: Results for subject PK. Each independent component is visualized with grouping and consistency information. The disks on the left show the spread of the cluster size and distance to other clusters. The four horizontal slices of the brain show the activation pattern from different heights. Also, the activation time-course is shown with the quantiles as different shades of gray.

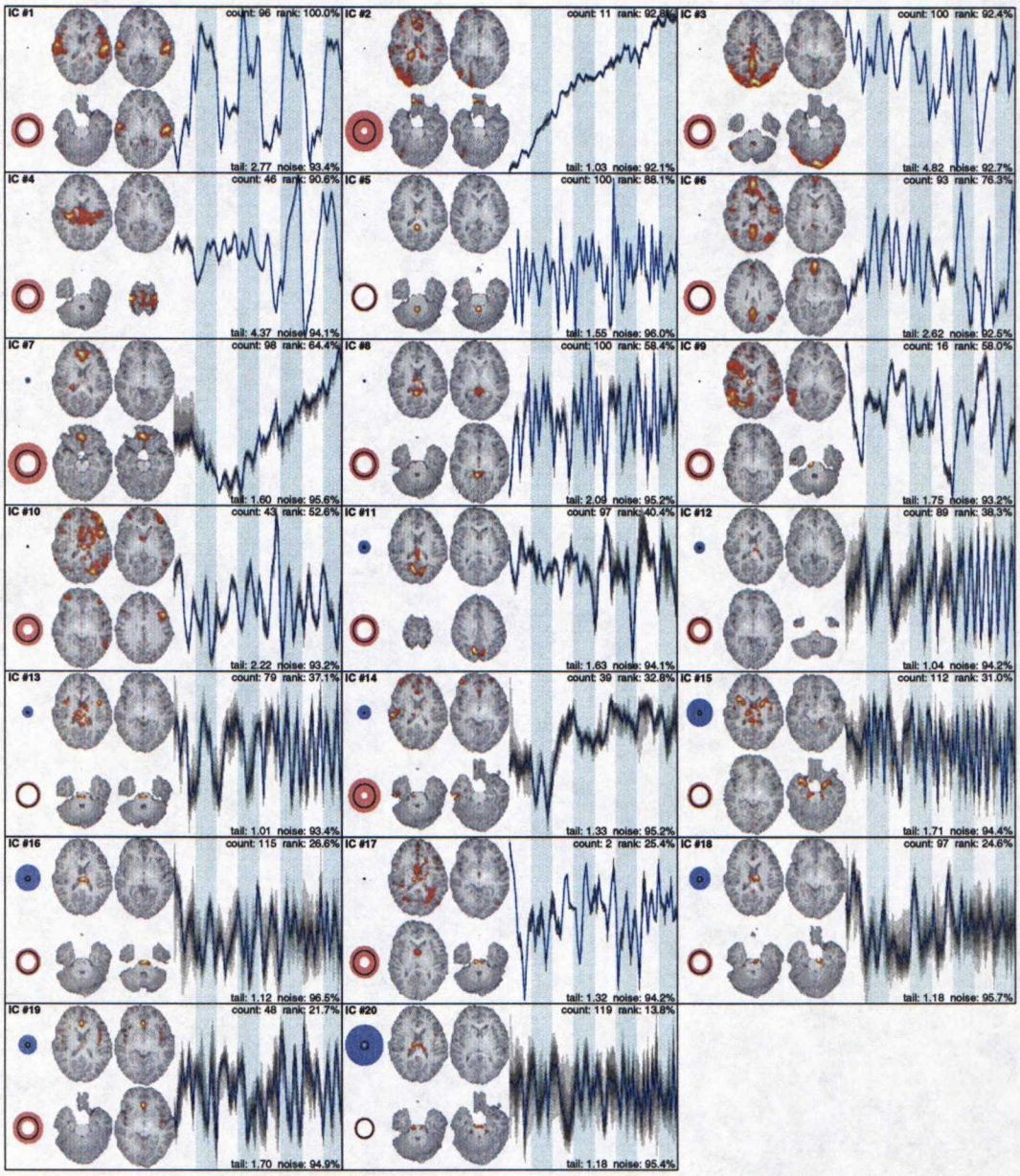


Figure B.9: Results for subject RS. Each independent component is visualized with grouping and consistency information. The disks on the left show the spread of the cluster size and distance to other clusters. The four horizontal slices of the brain show the activation pattern from different heights. Also, the activation time-course is shown with the quantiles as different shades of gray.

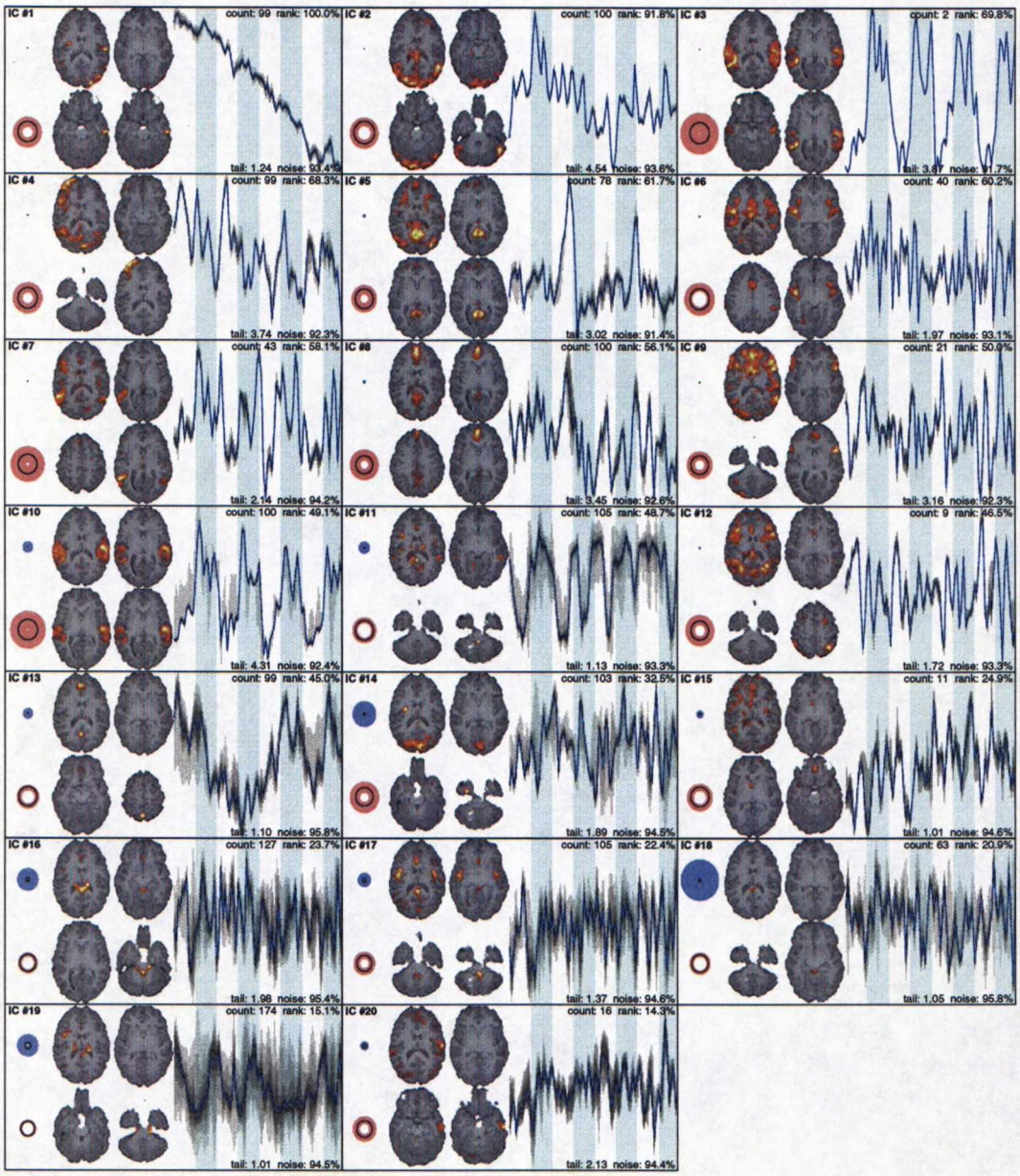


Figure B.10: Results for subject SN. Each independent component is visualized with grouping and consistency information. The disks on the left show the spread of the cluster size and distance to other clusters. The four horizontal slices of the brain show the activation pattern from different heights. Also, the activation time-course is shown with the quantiles as different shades of gray.

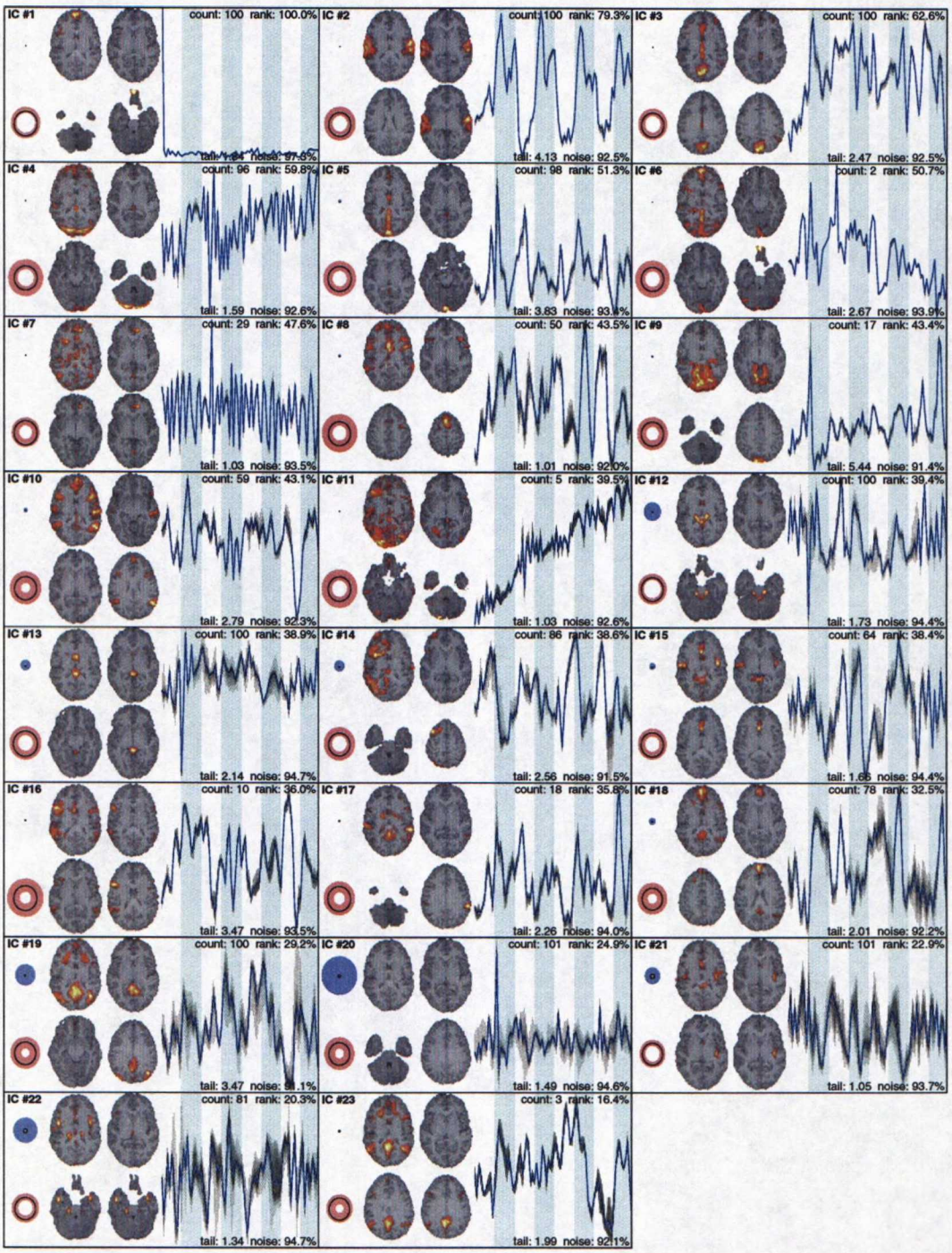


Figure B.11: Results for subject TL. Each independent component is visualized with grouping and consistency information. The disks on the left show the spread of the cluster size and distance to other clusters. The four horizontal slices of the brain show the activation pattern from different heights. Also, the activation time-course is shown with the quantiles as different shades of gray.

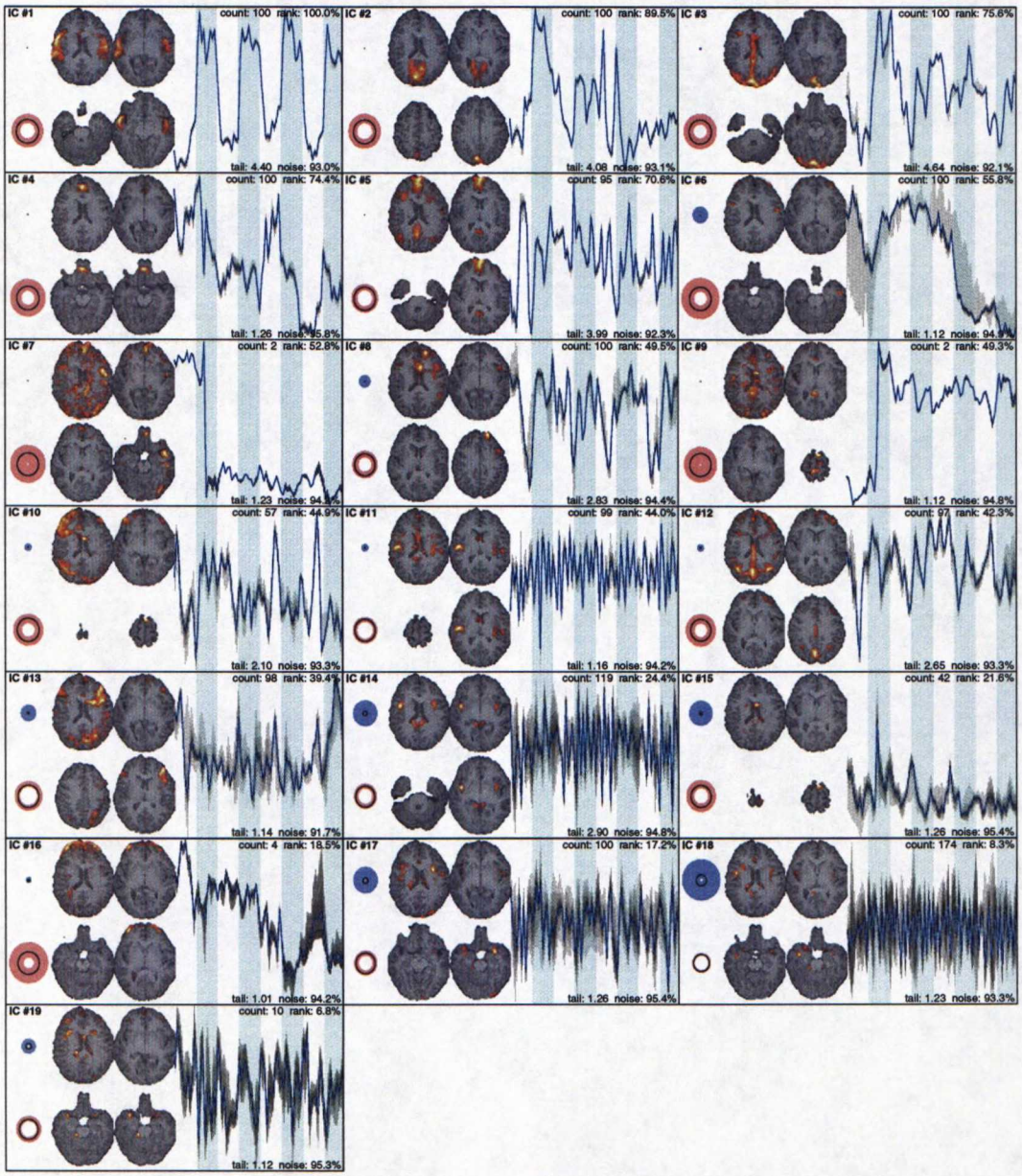


Figure B.12: Results for subject TP. Each independent component is visualized with grouping and consistency information. The disks on the left show the spread of the cluster size and distance to other clusters. The four horizontal slices of the brain show the activation pattern from different heights. Also, the activation time-course is shown with the quantiles as different shades of gray.

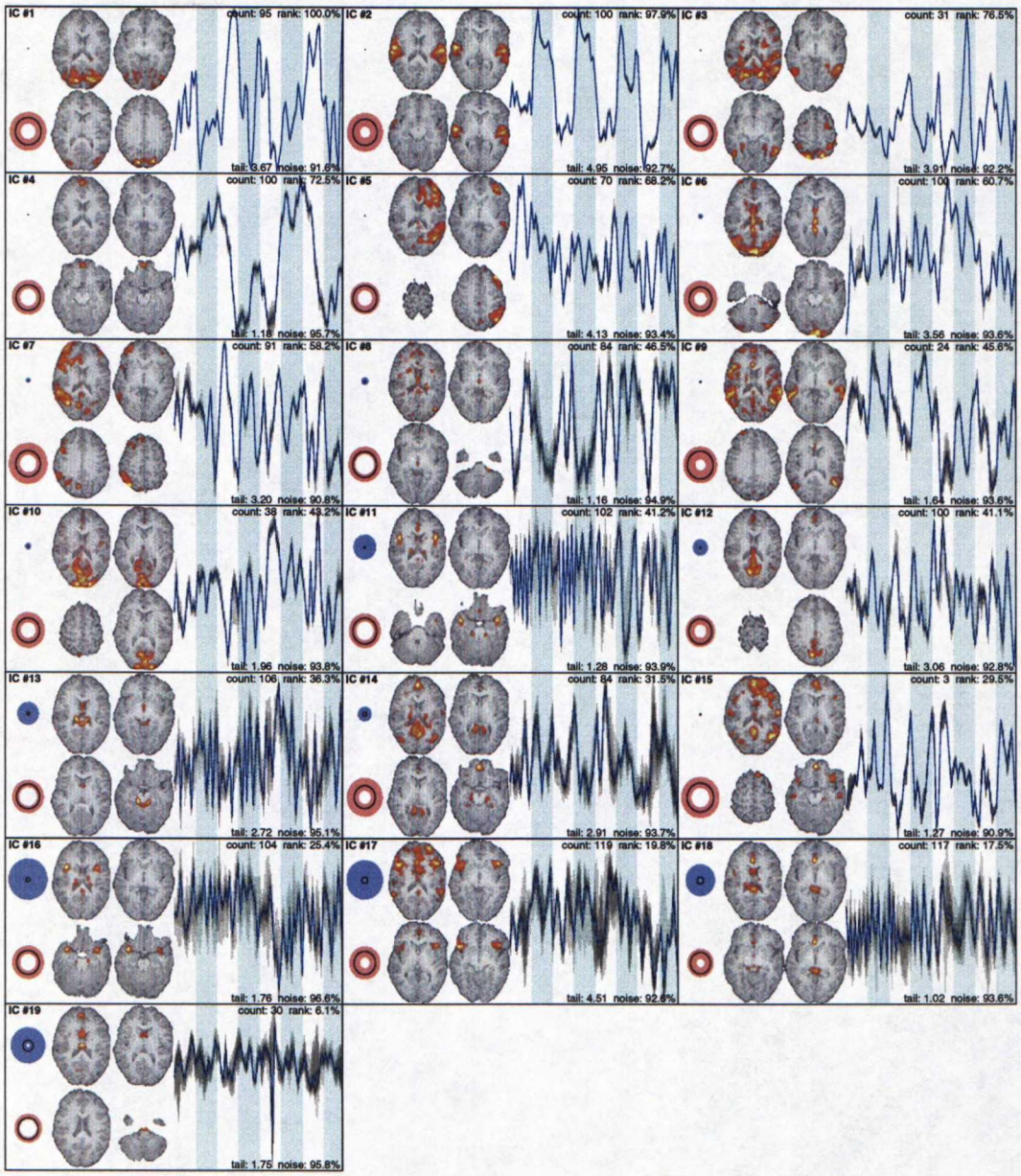


Figure B.13: Results for subject TT. Each independent component is visualized with grouping and consistency information. The disks on the left show the spread of the cluster size and distance to other clusters. The four horizontal slices of the brain show the activation pattern from different heights. Also, the activation time-course is shown with the quantiles as different shades of gray.

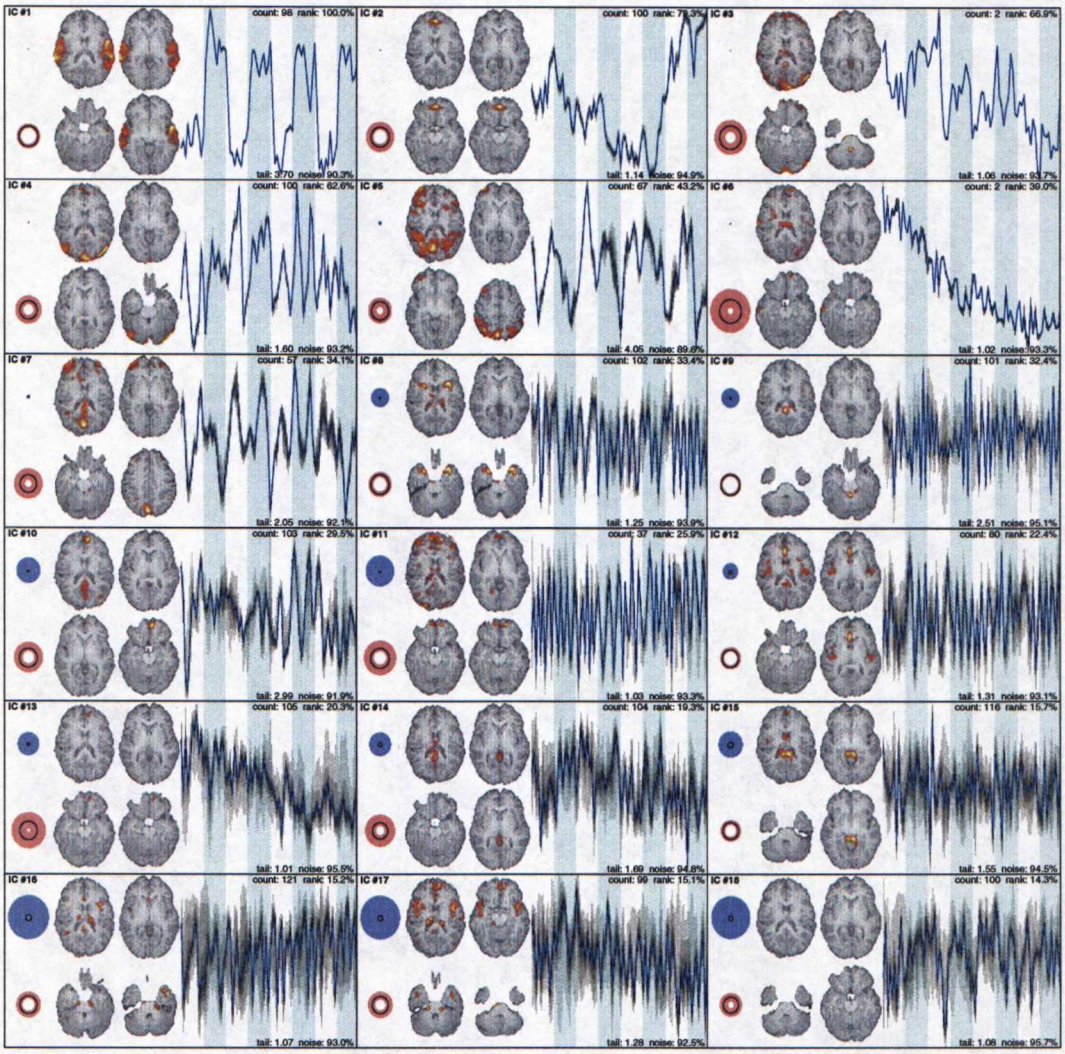


Figure B.14: Results for subject UL. Each independent component is visualized with grouping and consistency information. The disks on the left show the spread of the cluster size and distance to other clusters. The four horizontal slices of the brain show the activation pattern from different heights. Also, the activation time-course is shown with the quantiles as different shades of gray.

Bibliography

S.-I. Amari, A. Cichocki, and H. H. Yang. A new learning algorithm for blind signal separation. In *Proceedings of the 1995 Conference on Advances in Neural Information Processing Systems (NIPS 1995)*, volume 8, pages 757–763, Denver, CO, November 1995. 17

N. C. Andreasen. Evaluation of Brain Imaging Techniques in Mental Illness. *Annual Review of Medicine*, 39:335–345, February 1988. 8

S. R. Arnott, M. A. Binns, C. L. Grady, and C. Alain. Assessing the auditory dual-pathway model in humans. *NeuroImage*, 22(1):401–408, May 2004. 7

I. Bankman. *Handbook of Medical Imaging: Processing and Analysis.* Elsevier Science & Technology Books, Burlington, MA, 1st edition, October 2000. 7

A. Bartels and S. Zeki. The chronoarchitecture of the human brain — natural viewing conditions reveal a time-based anatomy of the brain. *NeuroImage*, 22(1):419–433, May 2004. 7

A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159, November 1995. 17

A. Belouchrani, K. A. Meraim, J.-F. Cardoso, and E. Moulines. Second-Order Blind Separation of Correlated Sources. In *Proceedings of the 1993 International Conference on Digital Signal Processing (DSP 1993)*, pages 346–351, Nicosia, Cyprus, July 1993. 16

V. D. Calhoun, T. Adali, L. K. Hansen, J. Larsen, and J. J. Pekar. ICA of functional MRI Data: An Overview. In *Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003)*, pages 281–288, Nara, Japan, April 2003. 23

V. D. Calhoun, T. Adali, V. B. McGinty, J. J. Pekar, T. D. Watson, and G. D. Pearlson. fMRI Activation in a Visual-Perception Task: Network of Areas

Detected Using the General Linear Model and Independent Components Analysis. *NeuroImage*, 14(5):1080–1088, November 2001a. 7, 54

V. D. Calhoun, T. Adali, G. D. Pearlson, and J. J. Pekar. A method for making group inferences from functional MRI data using independent component analysis. *Human Brain Mapping*, 14(3):140–151, August 2001b. 54

J.-F. Cardoso. Eigen-structure of the fourth-order cumulant tensor with application to the blind source separation problem. In *Proceedings of the 1990 International Conference on Acoustics, Speech, and Signal Processing (ICASSP 1990)*, volume 5, pages 2655–2658, Albuquerque, NM, April 1990. 16

J.-F. Cardoso. Dependence, Correlation and Gaussianity in Independent Component Analysis. *Journal of Machine Learning Research*, 4:1177–1203, December 2003. 58

P. Comon. Independent component analysis, A new concept? *Signal Processing*, 36(3):287–314, April 1994. 17

J.-R. Duann, T.-P. Jung, S. Makeig, and T. J. Sejnowski. Consistency of Infomax ICA Decomposition of Functional Brain Imaging Data. In *Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003)*, pages 289–294, Nara, Japan, April 2003. 28

FastICA. FastICA MATLAB Package, 1998. URL `http://www.cis.hut.fi/research/ica/fastica`. 20, 26, 61

E. Formisano, F. Esposito, N. Kriegeskorte, G. Tedeschi, F. D. Di Salle, and R. Goebel. Spatial independent component analysis of functional magnetic resonance imaging time-series: characterization of the cortical components. *Neurocomputing*, 49(1–4):241–254, December 2002. 33

O. Friman, M. Borga, P. Lundberg, and H. Knutsson. Detection and detrending in fMRI data analysis. *NeuroImage*, 22(2):645–655, May 2004. 54

T. Gautama, D. P. Mandic, and M. M. Van Hulle. A Novel Method for Determining the Nature of Time Series. *IEEE Transactions on Biomedical Engineering*, 51(5):728–736, May 2004. 33

S. Haykin. *Neural Networks: A Comprehensive Foundation*. Pearson/Prentice Hall, Upper Saddle River, NJ, 2nd edition, July 1998. 24, 55

J. Himberg, A. Hyvärinen, and F. Esposito. Validating the independent components of neuroimaging time series via clustering and visualization. *NeuroImage*, 22(3):1214–1222, July 2004. 31

S. A. Huettel, A. W. Song, and G. McCarthy. *Functional Magnetic Resonance Imaging*. Sinauer Associates, Sunderland, MA, 1st edition, April 2004. 10

A. Hyvärinen. Fast and Robust Fixed-Point Algorithms for Independent Component Analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, May 1999. 17, 20

A. Hyvärinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. Wiley-Interscience, New York, NY, 1st edition, May 2001. 17

A. Hyvärinen and E. Oja. A Fast Fixed-Point Algorithm for Independent Component Analysis. *Neural Computation*, 9(7):1483–1492, October 1997. 21, 61

A. Hyvärinen and E. Oja. Independent component analysis: algorithms and applications. *Neural Networks*, 13(4–5):411–430, June 2000. 18, 58, 61

Icasso. Icasso MATLAB Package, 2003. URL `http://www.cis.hut.fi/jhimberg/icasso`. 31

I. T. Jolliffe. *Principal Component Analysis*. Springer-Verlag, New York, NY, 2nd edition, October 2002. 19, 55

T.-P. Jung, S. Makeig, M. J. McKeown, A. J. Bell, T.-W. Lee, and T. J. Sejnowski. Imaging Brain Dynamics Using Independent Component Analysis. *Proceedings of the IEEE*, 89(7):1107–1122, July 2001. 22

C. Jutten and J. Herault. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24(1): 1–10, July 1991. 16, 17

J. W. Kalat. *Biological Psychology*. Thomson/Wadsworth, New York, NY, 8th edition, July 2003. 4

V. Kiviniemi, J.-H. Kantola, J. Jauhiainen, A Hyvärinen, and O. Tervonen. Independent component analysis of nondeterministic fMRI signal sources. *NeuroImage*, 19(2):253–250, June 2003. 22

M. J. McKeown, L. K. Hansen, and T. J. Sejnowski. Independent Component Analysis of functional MRI: What Is Signal and What Is Noise? *Current Opinion in Neurobiology*, 13(5):620–629, October 2003. 23

M. J. McKeown, S. Makeig, G. G. Brown, T. P. Jung, S. S. Kindermann, A. J. Bell, and T. J. Sejnowski. Analysis of fMRI Data by Blind Separation Into Independent Spatial Components. *Human Brain Mapping*, 6(3):160–188, August 1998. 22

M. J. McKeown and T. J. Sejnowski. Independent Component Analysis of fMRI Data: Examining the Assumptions. *Human Brain Mapping*, 6(5–6): 368–372, December 1998. 23, 25

R. L. McNamee and N. A. Lazar. Assessing the sensitivity of fMRI group maps. *NeuroImage*, 22(2):920–931, May 2004. 54

J. H. Meek, C. E. Elwell, M. J. Khan, J. Romaya, J. S. Wyatt, D. T. Delpy, and S. Zeki. Regional changes in cerebral haemodynamics as a result of a visual stimulus measured by near infrared spectroscopy. In *Proceedings of the Royal Society in Biological Sciences*, volume 261, pages 351–356, London, UK, November 1995. 9

F. Meinecke, A. Ziehe, M. Kawanabe, and K.-R. Müller. A Resampling Approach to Estimate the Stability of One-Dimensional or Multidimensional Independent Components. *IEEE Transactions on Biomedical Engineering*, 49(12):1514–1525, December 2002. 26

C. T. Moonen, P. C. van Zijl, J. A. Frank, D. Le Bihan, and E. D. Becker. Functional magnetic resonance imaging in medicine and physiology. *Science*, 250(4977):53–61, October 1990. 8

E. Niedermeyer and F. L. da Silva. *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields.* Lippincott, Williams & Wilkins, Baltimore, MD, 5th edition, November 2004. 8

S. Ogawa, D. W. Tank, R. Menon, J. M. Ellermann, S. Kim, H. Merkle, and K. Ugurbil. Intrinsic Signal Changes Accompanying Sensory Stimulation: Functional Brain Mapping with Magnetic Resonance Imaging. *Proceedings of the National Academy of Sciences*, 89(13):51–55, July 1992. 12

J. Särelä and H. Valpola. Denoising Source Separation. *Journal of Machine Learning Research*, 6:(In Press), March 2005. 34, 54

J. Särelä and R. Vigário. Overlearning in Marginal Distribution-Based ICA: Analysis and Solutions. *Journal of Machine Learning Research*, 4:1447–1469, December 2003. 22

SPM. SPM MATLAB Package, 1999. URL `http://www.fil.ion.ucl.ac.uk/spm`. 12, 14, 43

J. V. Stone. *Independent Component Analysis : A Tutorial Introduction.* MIT Press/Bradford Books, Cambridge, MA, 1st edition, September 2004. 17

J. Talairach and P. Tournoux. *Co-Planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System — An Approach to Cerebral Imaging.* Thieme Medical Publishers, New York, NY, 1st edition, January 1988. 41

K. J. Worsley and K. J. Friston. Analysis of fMRI Time-Series Revisited — Again. *NeuroImage*, 2(3):173–235, September 1995. 12, 14, 43

J. Ylipaavalniemi and R. Vigário. Analysis of Auditory fMRI Recordings via ICA: A Study on Consistency. In *Proceedings of the 2004 International Joint Conference on Neural Networks (IJCNN 2004)*, volume 1, pages 249–254, Budapest, Hungary, July 2004. 2, 33

A. Ziehe and K.-R. Müller. TDSEP — An Effective Algrithm for Blind Separation Using Time Structure. In *Proceedings of the 8th International Conference on Artificial Neural Networks (ICANN 1998)*, volume 8, pages 675–680, Skövde, Sweden, September 1998. 16